NASA Conference Publication 3132

Space Communications Technology Conference Onboard Processing and Switching

Proceedings of a Conference held in
Cleveland, Ohio
November 12–14, 1991

NASA

Space Communications Technology Conference Onboard Processing and Switching

Proceedings of a Conference held in
Cleveland, Ohio
November 12–14, 1991

Page intentionally left blank

FOREWORD

A surgeon in a rural health care facility transmits high-resolution video and imaging data from a procedure in progress and receives life-saving assistance from experts at a major metropolitan hospital... A consumer previews the offerings of competing suppliers and discusses features with an on-line representative before placing an order electronically from the personal computer-based terminal in his home... A scientist interacts with her spaceborne experiment in near-real time and modifies the test conditions to enhance the observation of an unexpected result... A university simultaneously communicates with teams of investigators on multiple research expeditions in remote locations...

These are some examples of future services enabled by the use of advanced signal processing and switching technology onboard the next generation of communications satellites. In order for these visions to become a reality, continued advances in space- and ground-based electronics and systems technologies are needed. For this reason, the National Aeronautics and Space Administration's Office of Aeronautics, Exploration, and Technology sponsors a program in onboard processing and switching technology development conducted by the Space Electronics Division at Lewis Research Center in Cleveland, Ohio.

The Space Electronics Division periodically conducts industry briefings, workshops, and conferences to bring together participants from industry, academia, and government to survey the recent advances in space communications and electronics technology and to focus future development activities. The second **NASA Space Communications Technology Conference** held in Cleveland, Ohio, on November 12-14, 1991, focused on recent developments in **Onboard Processing and Switching** and addressed the challenges of inserting this technology into future communications satellite systems.

The architectures for future satellite networks will draw on distributed processing with low-cost Earth terminals, power-efficient access techniques, and intelligent onboard resource control to provide new classes of circuit- and packet-switched services compatible with emerging national and international communications protocols at a cost competitive with terrestrial alternatives. Reconfigurable, fault-tolerant, electronics subsystems onboard the spacecraft will be required to support changing traffic and service demands and to ensure maximum utilization of available resources over an extended life. Advances in multichannel communications electronics will enable access from hundreds of thousands of users with cost-efficient terminals. Digital modulation and coding techniques will reduce the power and bandwidth requirements of future satellite systems while compensating for intrinsic distortions. This range of topics is addressed in the conference proceedings that follow.

Clearly, the technical and economic challenges of applying onboard processing and switching technology to future satellite networks are large, but they are not insurmountable. Proposed concepts suggest that improvements in bandwidth and power efficiencies by a factor of two or more are possible and that flexible, robust onboard subsystems may reduce power consumption and mass by an order of magnitude. The challenge will be to develop these technologies to a level of maturity and to demonstrate them in an appropriate systems environment to promote their adoption for operational use. Meeting this challenge requires that we effectively share and build upon the accomplishments of the many capable individuals and organizations that are

developing the technologies necessary to provide the next significant advancement to commercial and government communications.

James M. Budinger

Conference Chairperson
Deputy Chief, Digital Systems Technology Branch
NASA Lewis Research Center, Cleveland, Ohio

CONTENTS

Information Switching and Routing

Modulation and Coding

---

[1]These papers were not available in time for publication in this proceedings. They will be presented in a companion publication.

ON-BOARD PROCESSING ARCHITECTURES
FOR SATELLITE B-ISDN SERVICES[1]

Thomas Inukai, Dong-Jye Shyy, and Faris Faris
COMSAT Laboratories
Clarksburg, Maryland 20871

SUMMARY

This paper addresses on-board baseband processing architectures for future satellite broadband integrated services digital networks (B-ISDNs). To assess the feasibility of implementing satellite B-ISDN services, critical design issues, such as B-ISDN traffic characteristics, transmission link design, and a trade-off between on-board circuit and fast packet switching, are analyzed. Examples of the two types of switching mechanisms and potential on-board network control functions are presented. A sample network architecture is also included to illustrate a potential on-board processing system.

1. INTRODUCTION

The B-ISDN will likely play a major role in the future telecommunications networking for providing high speed integrated services to network users such as broadband video-telephony, broadband videoconference, high volume file transfer, high speed telefax, high definition TV (HDTV), and broadband videotext. It is envisaged that this technology will have an impact on satellite communications. This paper addresses some of the key design issues and alternate on-board switching architectures for a satellite-based B-ISDN.

The paper presents examples of B-ISDN traffic characteristics and identification of potential satellite applications of B-ISDN services. One of the primary concerns in designing a satellite B-ISDN system is the design of a transmission link that can support very high data rates ranging from 155 Mbit/s to over 1 Gbit/s with an availability that is comparable to terrestrial-based services. A summary of link analysis for Ku- and Ka-bands is included in the paper.

One of the most critical design issues for on-board processing satellites is the selection of an on-board baseband switching architecture. Circuit switching and fast packet switching are probably the two most common architectures. A trade-off analysis is performed for these architectures on the capability of handling circuit- and packet-switched traffic and the impact of traffic reconfiguration. Several switch structures for the two types of switching are also illustrated.

Potential satellite network architectures for B-ISDN include baseband-switched TDMA (SS-TDMA), TDMA up-link/TDM down-link, TDM up-link/TDM down-link, and hopping beam TDMA. SS-TDMA requires simpler on-board hardware and provides flexible connectivity among the same type of earth stations. TDMA up-link/TDM down-link allows the use of shared up-link capacity by multiple earth stations, while optimizing down-link transmission on a single carrier. It will require an on-board baseband processor for rate conversion and interconnection of user traffic. TDM up-link/TDM down-link is particularly suited for trunking applications and circuit-switched B-ISDN traffic. As in the previous case, an on-board processor provides rate conversion and connectivity. Thin-route traffic can be efficiently carried by hopping beam TDMA to dynamically

1

allocate the necessary capacity to different dwell areas. The paper presents a sample network architecture with TDMA up-link and TDM downlink and on-board fast packet switching.

2. SATELLITE APPLICATIONS

The B-ISDN supports a wide variety of communications services. Although the B-ISDN is being developed primarily for terrestrial networking using fiber optic cables, satellites are well positioned in complementing terrestrial-based B-ISDN services. The satellite-based system has inherent capabilities of providing multipoint/broadcast transmission, connectivity between any two points within a beam coverage, quick reallocation of space segment capacity, and distance-insensitive cost. Use of multiple spot beams and on-board rate conversion/switching will provide a larger capacity, additional flexibility, and lower user terminal cost. The satellite B-ISDN system can be used for direct (mesh) interconnection among users (UNI-UNI), interconnection between a user and a switching node (UNI-NNI), and interconnection between switching nodes (NNI-NNI). Figure 1 illustrates these options.
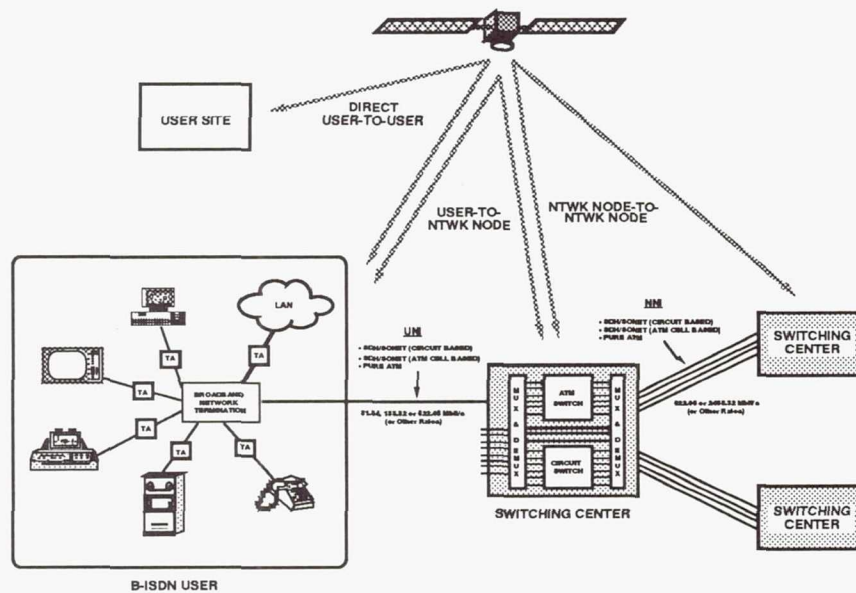
Figure 1. A satellite B-ISDN can provide direct interconnection between users, switching centers, and a user and a switching center.

A variety of B-ISDN services can be supported by a satellite network. Table 1 depicts examples of satellite B-ISDN services and their typical traffic characteristics. The B-ISDN interface rates are 155.52 Mbit/s (51.84 Mbit/s for SONET STS-1) or higher, but their information rates can be as small as 64 kbit/s. Traffic can be extremely bursty such as in LAN/MAN interconnection, or can be a steady flow of high speed data as in video program distribution. A satellite network must be flexible to accommodate a wide range of transmission rates and different degrees of burstiness. In addition, Asynchronous Transfer Mode (ATM) is a packet-based transmission mechanism that supports connection-oriented as well as connectionless services at various bit rates. Thus, a satellite system design must take into account these new environments that are significantly different from the traditional circuit-oriented system.

2

Table 1. Traffic Characteristics of Satellite B-ISDN

SAMPLE APPLICATION	BIT RATE RANGE (Mbit/s)	BURSTINESS	SOURCE TRAFFIC TYPE
Video Distribution	45 - 140	Low	Circuit
Live Newscast/Broadcast	20 - 45	Low	Circuit
Science Data Distribution	30 - 300	Medium	Circuit/Packet
Supercomputer Networking	30 - 1000	High	Packet (Circuit)
Private Networking	0.1 - 100	High	Packet (Circuit)
Trunking	52 - 2500	Low	Circuit
Emergency Communications	2 - 45	Low to Medium	Circuit (Packet)
Thin-Route Networking	0.1 - 10	Medium to High	Packet (Circuit)

3. TRANSMISSION LINK DESIGN

As seen in the previous section, the satellite B-ISDN network must support a wide range of bit rates from as low as 64 kbit/s to over 1 Gbit/s. A link analysis was performed for typical Ku- and Ka-band satellites to determine typical bit rates that can be supported by various earth station sizes (HPA and antenna). A Ku-band spot beam satellite assumes INTELSAT-VII or typical domestic satellite parameters with a G/T of 5 dB/K and an EIRP of 50 dBw, and a Ka-band satellite has a G/T of 20 dB/K and an EIRP of 60 dBw (ACTS parameters). It is also assumed that these satellites perform on-board regeneration as well as FEC decoding and recoding. The results of link analysis are shown in Table 2.

Table 2. Link Analysis Results (QPSK Modulation, BER=10-8)

FREQ. BAND.	ANTENNA DIAMETER (m)	HPA SIZE (watts)	BIT RATE (Mbit/s)	FEC CODING RATE	UPLINK MARGIN (dB)	DOWNLINK MARGIN (dB)
	12.0	100	620	0.75	10.8	12.5
	9.0	100	620	0.75	8.3	10.0
	7.5	100	310	0.5	10.8	12.5
Ku	5.0	50	155	0.5	7.3	12.0
	2.4	50	50	0.5	5.8	10.5
	2.4	20	10	0.5	8.9	10.5
	1.2	20	10	0.5	2.8	4.5
	12.0	100	1240	0.75	22.8	15.8
	9.0	100	1240	0.75	20.3	13.3
	5.0	50	620	0.5	16.3	12.3
Ka	5.0	50	310	0.875	17.3	13.3
	2.4	50	155	0.5	15.9	11.9
	2.4	20	50	0.5	16.9	16.9
	1.2	20	10	0.5	17.8	10.8

The results indicate that an information rate of over 1 Gbit/s can be supported by a large Ka-band earth station, and small VSAT class terminals can be used for bit rates of up to 155 Mbit/s. A relaxed BER requirement, e.g. 10^{-6}, will result in a greater margin, a smaller antenna/HPA size, or a higher bit rate. If an additional margin or a lower threshold BER, e.g. 10^{-11}, is needed, a high-rate outer code, such as a Read-Solomon code, may be used with a slight increase in the required bandwidth.

4. CIRCUIT OR FAST PACKET SWITCHING?

On-board baseband switching provides interconnection of user earth stations operating at different bit rates and access schemes. There are two types of switching architectures that can be used for baseband switch implementation: (a) circuit switching and (b) fast packet switching. Circuit switching uniquely maps uplink time slots to downlink time slots and guarantees their connections until a controller deallocates the assigned slots. In fast packet switching, data are packetized into a fixed format with a routing header, and packets from earth stations are routed to the destination beams (or carriers) according to header information[2]. No fixed mapping exists between uplink slots and downlink slots. However, there is a potential on-board buffer overflow problem. A comparison of the two switching architectures is shown in Table 3.

Table 3. Comparison of Circuit and Fast Packet Switching

SWITCHING ARCHITECTURE	CIRCUIT-SWITCHED TRAFFIC	PACKET-SWITCHED TRAFFIC	TRAFFIC RECONFIGURATION
CIRCUIT SWITCHING	• Efficient Bandwidth Utilization	• Very Inefficient Bandwidth Utilization • Inflexible Connectivity	• Reprogramming of On-Board Switch Control Memories • Reconfiguration of Earth Station Time/Frequency Plans for Each Circuit Setup • Difficult to Implement Autonomous Private Networks
FAST PACKET SWITCHING	• Can Accommodate Circuit-Switched Traffic • Somewhat Higher Overhead Due to Packet Headers	• On-Board Congestion May Occur	• Self-Routing • Does not Require Control Memory for Routing • Reconfiguration of Earth Station Time/Frequency Plans for Major Traffic Changes • Easy to Implement Autonomous Private Networks

From this table, the following conclusions can be made:

a. Select circuit switching for circuit-switched traffic without frequent traffic reconfiguration.

b. Select fast packet switching for packet-based traffic or circuit-switched traffic with frequent channel reconfiguration. Special consideration must be given to the on-board congestion problem.

The congestion problem for fast packet switching can be completely eliminated for circuit-switched traffic by allocating a desired capacity (on a call-by-call basis) for both uplink and downlink carriers. For packet-switched traffic, some form of flow/congestion control is needed to minimize on-board buffer overflow. The techniques include the following: (a) a dynamic allocation of fixed capacity from a transmit earth station to each downlink carrier, (b) call admission control at the earth station, (c) on-board capacity allocation based on current queue status, (d) feedback control, and (e) a combination of these. The goal of flow/congestion control is to achieve a certain packet

---

[2]On-Board Processing Satellite Network Architecture and Control Study, Final Report, NASA Contract NAS3-24886, Prepared by COMSAT Laboratories, June 1987. Also, see a paper by S. J. Campanella and B. A. Pontano, "Future Switching Satellite," AIAA 12th International Communication Satellite Systems Conference, Virginia, pp. 264-273, March 13-17, 1988.

loss ratio within a satellite system such that no significant degradation results in the end-to-end performance. Some of these techniques have been successfully tested at COMSAT Laboratories.

The on-board switch selection described above is a general guideline based on a qualitative trade-off and should be carefully evaluated for specific applications, considering a total system capacity, traffic types, user traffic volume, and network connectivity.

5. ON-BOARD BASEBAND SWITCHING ARCHITECTURES

The simplest form of a baseband circuit switch consists of a space switch, such as the one used in the ITALSAT system, to provide SS-TDMA operation. The space switch allows dynamic interconnection among uplink and downlink spot beams according to the switch state configurations stored in the on-board control memory. It provides high speed interconnection most efficiently with a minimum amount of hardware and is also capable of providing multicast connectivity. A typical B-ISDN application of SS-TDMA includes HDTV program distribution, trunking, and cable restoration.

More flexible circuit switching to provide rate conversion and channel demultiplexing/multiplexing is achieved by the use of on-board data memory. Examples of this type of switch are a common memory (T-stage) switch, a distributed output memory with a parallel data bus (S-T), and a distributed input/output memory with a space switch (T-S-T). Figure 2 illustrates these switch structures. Also shown in the figure is a switch structure using a high speed fiber optic ring. The common memory structure possibly requires the least amount of hardware, but its capacity is limited by the memory access speed. The capacity of a distributed output memory structure depends on the speed of the parallel bus, and a careful design is required to avoid signal interference among the bus lines. The T-S-T structure is modular in design and can accommodate a larger capacity, but it may require a larger buffer size. The fiber optic bus is also modular in design, embodies a fault-tolerant structure, and can achieve a high throughput. Selection of a particular switch structure must consider the throughput requirement, a redundancy structure, and mass/power requirements.
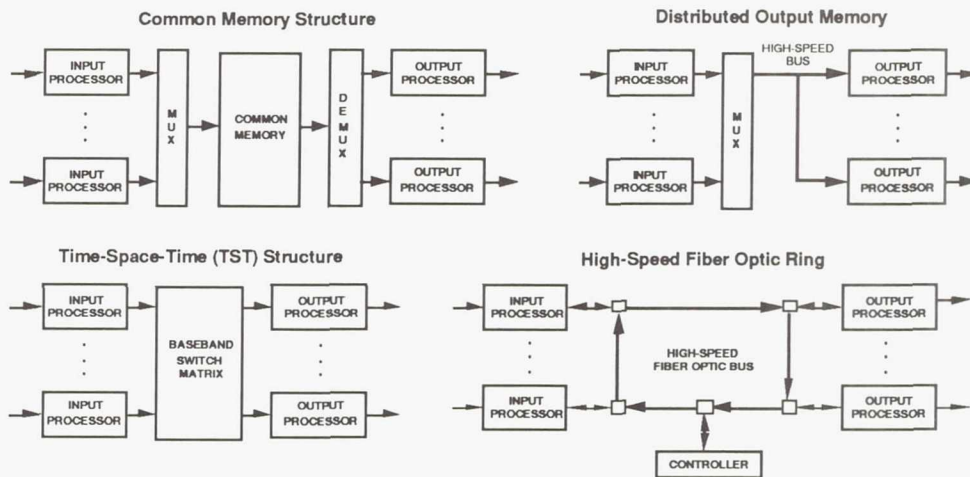
Figure 2. Examples of Circuit Switch

Fast packet switching allows dynamic routing of small packets to desired destination beams solely based on their routing headers. The routing function is performed with a hardware-oriented self-routing architecture. In fast packet switching, the bandwidth assigned to each connection can be instantaneously varied, which circuit switching cannot achieve. The interconnection network of the packet switch decides the performance of the switch such as delay and throughput. The circuit

switch structures described above, except T-S-T, can also be used for fast packet switching, but the switch throughput is limited by the speed of the shared memory or the shared medium.

The space-division network can set up multiple connections from different input ports to different output ports. Since packet transfer is done in parallel, the switching network speed is reduced compared with the shared-memory and shared-medium schemes. The other advantage is that routing control can be distributed over the switching fabric. Examples of space division switching networks are shown in Figure 3.
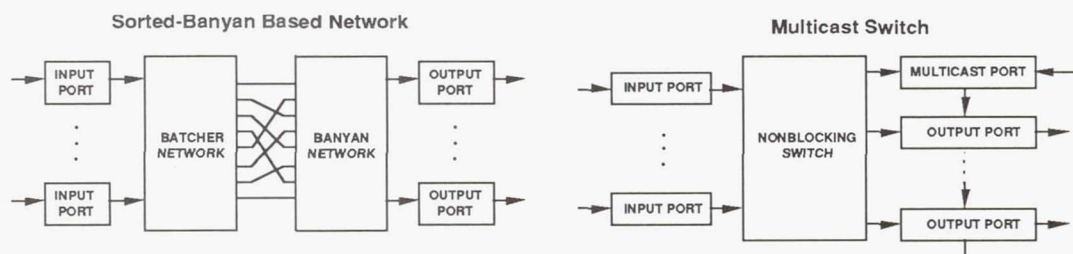
Figure 3. Examples of Fast Packet Switch Using a Space Division Switching Network

The space-division switching network can be implemented either in electronics, optoelectronics, or optics. The maximum data rate of a switch is a function of the device technology and the delays of the links. The VLSI technologies such as GaAs can support a data rate of up to several thousand Mbit/s. Other fast VLSI technologies are ECL and high speed CMOS.

Fast packet switching is particularly suited for providing dynamic interconnection of users with bursty traffic. Its application is not limited to only ATM users but includes lower speed traffic users with N-ISDN circuit-switched or packet-switched (X.25/frame relay) traffic.

In the conventional transparent satellites, network control functions, such as allocation of space and ground segments, demand assignment processing, and earth station monitor and control are performed by a ground-based network control station (NCS) through orderwire communication channels. The NCS is often equipped with sophisticated computer systems to perform real-time and batch processing to support dynamic network operation. Implementation of network control functions on board the satellite not only simplifies ground-based network control, but also provides a more efficient, reliable, and simpler interface (i.e., direct error-free connection between the on-board processor and the controller) and a lower ground segment cost. Some of the key on-board network controller (OBNC) functions are shown in Figure 4.
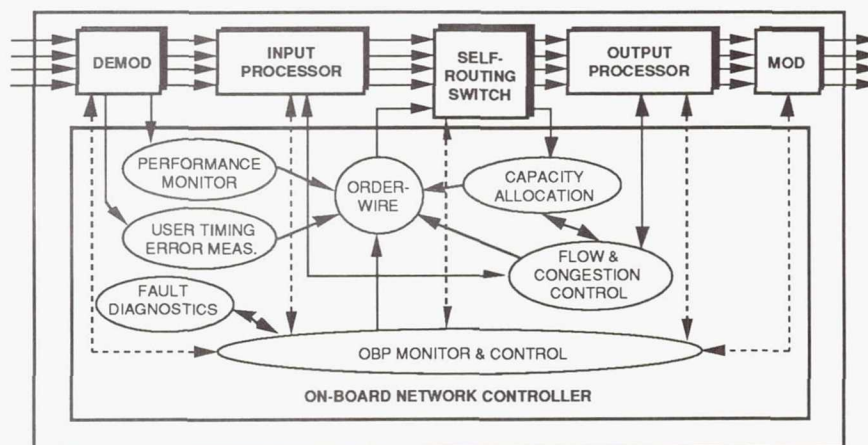
Figure 4. On-Board Network Controller Functions

To minimize the complexity of implementing these functions, expert systems and/or neural network technology may be fully utilized.

7. SAMPLE NETWORK ARCHITECTURE

A sample network architecture for an advanced on-board processing system is illustrated in Figure 5. The system consists of 15 Ka-band steerable beams and 13 Ku-band fixed spot beams, both covering the CONUS. The Ka-band beams can be parked on any of the 87 spot beam areas according to the beam traffic requirements. Each beam supports 155 Mbit/s TDMA transmission. The Ku-band system operates on an uplink TDMA bit rate of 5 Mbit/s and 10 carriers per beam and a single carrier TDM downlink at 50 Mbit/s. The Ku-band system is primarily used by low traffic users, for example, with N-ISDN interfaces. The two types of systems are interconnected by a 20 x 20 155-Mbit/s baseband switch matrix (BSM) and an 18 x 18 fast packet switch. The system capacities are respectively 2.3 Gbit/s and 650 Mbit/s for Ka- and Ku-bands, and a simple non-blocking switch structure shown in Figure 2 may be used for fast packet switch implementation.
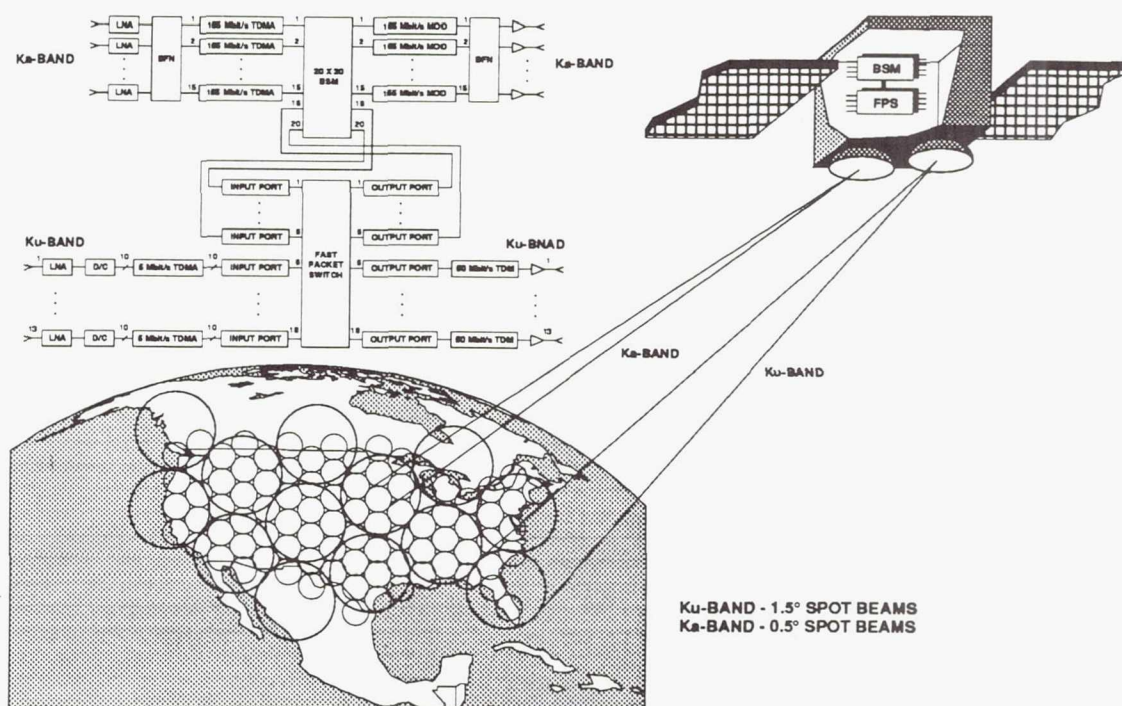
Figure 5. Sample Network Architecture

8. CONCLUSION

Satellite systems can provide B-ISDN services at bit rates of up to 1.24 Gbit/s with an antenna size of 9 m (Ka-band). However, user bit rates can be significantly smaller than this upper bound in most applications. A typical information rate may vary in the range of 100's kbit/s to 10's Mbit/s with occasional requirements at 52 Mbit/s and at up to 155 Mbit/s. In this scenario, cost-effective satellite integrated digital services can be provided to users with an antenna size of 1.8 - 2.4 m, and the use of fast packet switching on board the satellite will simplify overall network control and provide additional flexibility.

Page intentionally left blank

DESTINATION DIRECTED PACKET SWITCH ARCHITECTURE FOR A 30/20 GHz FDMA/TDM GEOSTATIONARY COMMUNICATION SATELLITE NETWORK

William D. Ivancic and Mary Jo Shalkhauser
NASA Lewis Research Center
Cleveland, Ohio

Abstract

This paper concentrates on a destination directed packet switching architecture for a 30/20 GHz FDMA/TDM geostationary satellite communications network. Critical subsystems and problem areas are identified and addressed.

Efforts have concentrated heavily on the space segment; however, the ground segment has been considered concurrently to ensure cost efficiency and realistic operational constraints.

Introduction

In the mid 1980's NASA began the Advanced Communication Technology Satellite (ACTS) program to develop a 30/20 GHz geostationary communication satellite to be launched in 1993. This satellite will open up the Ka-band frequency for commercial communications, develop multibeam and hopping-beam antennas, and demonstrate onboard processing technology. The ACTS system utilizes time division multiple access (TDMA) uplinks and time division multiplexed (TDM) downlinks. One of the drawback of TDMA uplinks are that the ground terminals are forced to transmit at a much higher data rate than there actual throughput rate. For example, in the ACTS system, a ground terminal wishing to transmit a single voice channel at 64 kbps would have to transmit at a burst rate of 27.5, 110, or 220 Mbps (ref. 1). This, in effect, drives the cost of the ground terminals up dramatically by requiring either substantially higher power transmitters or larger antennas, both which are major cost drivers in a low cost ground terminal. Realizing this, recent emphasis has been place on driving the cost of the ground terminals down. One way to accomplish this is to eliminate the need for high power transmitters on the ground by allowing the user to transmit at a lower data rate using a frequency division multiple access (FDMA) uplink architecture. TDM is chosen for the downlink transmission technique because the high power amplifier (HPA) can be operated at maximum power thereby increasing the downlink signal strength which, in turn, enables the use of very small aperture terminals or VSATs (ref. 2).

Currently, NASA envisions the need for meshed VSAT satellite communications systems for direct distribution of data to experimenters and direct control of space experiments. In the commercial arena, NASA envisions a need for low data rate, direct to the user communications services for data, voice, FAX, and video conferencing. Such a system would enhance current communications services and enable new services. For this type of satellite systems to exist, it must be cost competitive with terrestrial systems at the user level while enhancing the existing quality of service. The key to making this system cost competitive is to drive the cost of the ground terminals down and spread the cost of the satellite among tens or thousands of users. NASA has completed and is continuing to performed a number of studies on such communication systems (ref. 3-6).

Meshed VSAT satellite networks can be

implemented using either a circuit switched architecture, a packet switched architecture, or a combination of the two. Intuitively, it appears that a circuit switched network would be far simpler to implement; however, a packet switch has many potential advantages relative to circuit switching. Therefore, the Digital System Technology Branch at NASA LeRC is currently investigating a packet switched satellite network in order to identify the common subsystems of a circuit and packet switched network and to quantify the complexity of a packet switched network verses a circuit switch. This paper is a direct result of those studies.

The paper will describe the overall network requirements, the network architecture, the protocols and congestion control, and the individual subsystems of a destination directed packet switched geostationary satellite network for commercial communications.

Network Requirements

In order to begin designing the conceptual satellite architecture a list of salient requirements has to be created. The requirements follow: First, the system has to be economically viable and cost competitive with existing terrestrial telecommunication systems while enhancing existing services and adding new ones. Second, the system must provide voice, data, FAX, datagram, teleconferencing, and video communications services. In order to provide these services, the ground terminals will either transmit fixed length packets at 64 kbps or transmit continuously at 2.048 Mbps. It is envisioned that at 2.048 Mbps, the required service will be trunked continuous transmission circuits analogous to the present practice of leasing dedicated T1 circuits. Third, the system will be capable of point-to-point, multicast, and broadcast transmission. Multicast capability is a necessity in order to provide teleconferencing and video conferencing services. Broadcast transmission may not be necessary but is desirable. The requirement to communicate to every user in the system simultaneously (broadcast) verses only a select number of users within the system (multicast) is not readily apparent. Fourth, the satellite has to accommodate destination directed packets on a packet-by-packet basis. There does not appear to be any advantage to using packets versus a simple circuit switch through a satellite system unless they are destination directed -- the packet destination is contained in each header. For example, the use of packets to set up a virtual circuit simply adds the complexity of packet synchronization and processing to what is actually a circuit switch. Fifth, the satellite will not drop packets. Due to the long round-trip delay times to geostationary satellites, 250 msec, if packets are dropped, the window for requesting a retransmission and the data buffering involved becomes quite undesirable.

Network Architecture Description

The network consists of meshed VSATs operating at 30/20 GHz, transmitting through a processing satellite. Transmission is FDMA up and TDM down. There are eight uplink beams and eight downlink hopping beams covering CONUS. Each downlink beam has eight dwell locations. Associated with each uplink beam is a multi-channel demultiplexer/demodulator capable of demultiplexing and demodulating one thousand and twenty-four 64 kbps channels, a packet synchronizer, a decoder, and a MCDD-to-switch formatting buffer. Associated with each downlink is a switch-to-TDMA formatting buffer, an encoder, and a 150 Mbps burst modulator. The NxN

10
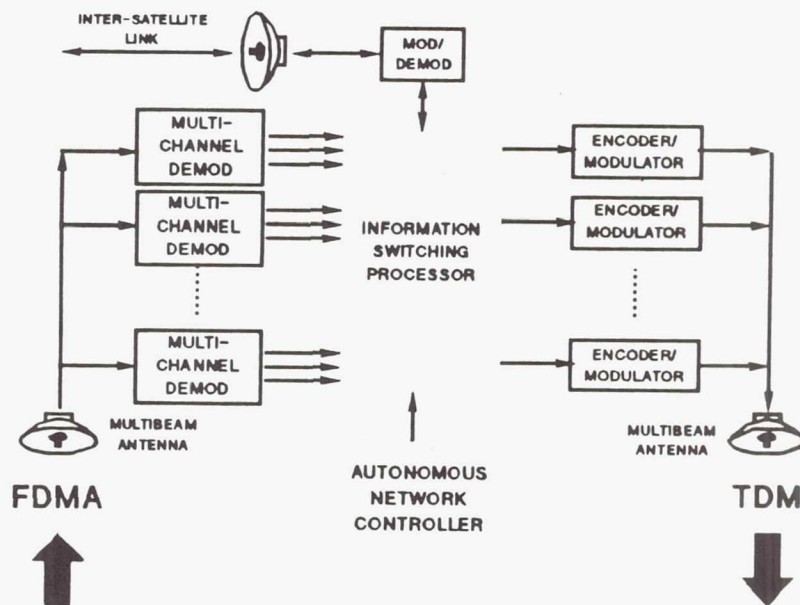
Figure 1   FDMA/TDM Network
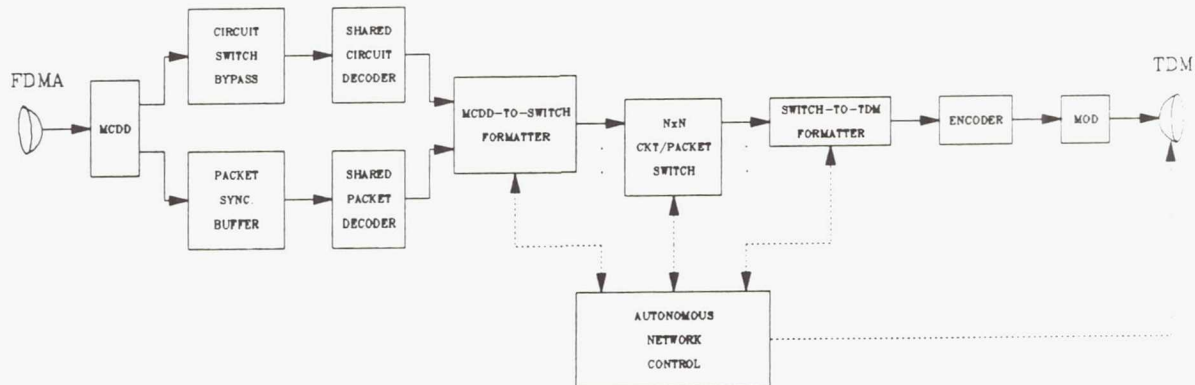
Figure 2   ISP Detailed Architecture

11

switch performs the spacial switching functions while the formatting buffers perform the temporal switching. The switching, routing, and congestion control are the responsibility of the autonomous network controller onboard the satellite. (fig. 1 and 2)

Many of the relative numbers used to establish the network size and data rates are taken from an architecture study perform by TRW. This report includes a complete link budget and hardware analysis in sufficient detail to estimate size, weight, and power requirements for the satellite and ground terminals (ref. 7).

Protocols

Initial Access

Initial access into the system would be via a reserved signalling channel. One channel for each multichannel demultiplexer would be reserved for requesting entry into the system. This channel would be set up in a slotted aloha format (ref. 8) and accept 64 kbps packets containing a request for a data transmission rate corresponding to either 64 kbps packet data transmission or 2.048 Mbps circuit transmission. Additional information that may be conveyed during initial access would be related to type of data being transmitted (voice, video, FAX, datagram, etc...) and the effective data throughput rate. This information may be useful to the autonomous network controller in order to anticipate and correct for congestion problems. Also, during initial access, the ground terminal will have to specifically request for multicast or broadcast services. This is necessary in order to verify that the downlink capacity can handle the request and to properly bill the user for these services. For broadcasts and multicasts, the satellite will have to be capable of duplicating the received message up to 64 times onboard the satellite and place that information in to correct downlink beam and dwell.

Upon reception of the initial access request, the satellite will respond via a downlink inband orderwire message as to request granted or denied and a corresponding frequency allocation.

Packet Formats

For 64 kbps packet transmission, the ground terminal will translate incoming data -- be it packets, voice, continuous data, etcetera -- into packets that are specific to the satellite network. The data packets are fixed length in order to simplify the onboard processing. All flow control, acknowledges, and buffering are performed at the ground terminal. There are six fields specific to the packet: synchronization, destination address, source address, control, information, and parity (fig 3).

The synchronization field is used to determine the start of the packet. This field is only necessary in an asynchronous packet network where no timing structure is overlaid on the transmitting portion of the ground terminals. The synchronization field has to be long enough to reduce the probability of a false detection without being so long as to dramatically increase the packet overhead. Presently, this field is 32 bits long.

The destination address field specifies the downlink destination which consists of the downlink beam and the dwell location within that beam. Twenty-six bits are reserved for this corresponding to the eight downlink beams, eight dwell locations within each beam, and 1024 possible ground terminals.

SYNC. HEADER	DESTINATION ADDRESS	SOURCE ADDRESS	CONTROL	INFORMATION	PARITY
1　　　　32	33　　　　58	59　　　　74	75	76	
32 BITS	26 BITS	16 BITS	1 BIT	TBD BITS	TBD BITS
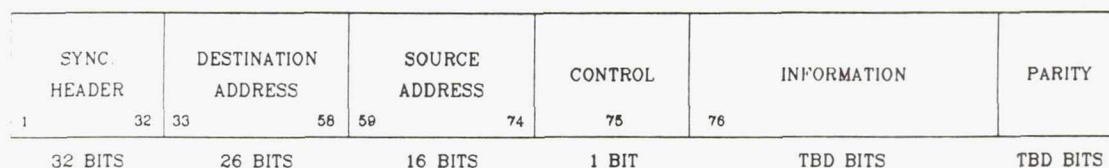
Figure 3 Packet Format

The source address field specifies the uplink source. Sixteen bits are reserved for this: three specify the uplink beam, ten specify the uplink frequency corresponding to the transmitting ground terminal; and three specify one of eight multiplexer input ports where up to eight active users may share one ground terminal simultaneously.

There is a one bit control field that indicates whether or not the packet contains useful information or is a dummy packet. Dummy packets are not passed on to the switch but are simply used by the demodulator to maintain lock.

The information field contains communications data that is being passed from ground terminal to ground terminal. This can be continuous transmission data such as voice or standard packets that have the satellite network packet structure overlaid. The length of the packet has not been determined at this time. The tradeoff on packet length is between improved packet efficiency and increased onboard storage. The longer the packet the greater the packet efficiency due to a reduction in the overhead-to-information ratio; however, the longer the packet the greater the onboard storage requirements.

The address, control, and information fields are error correction encoded and that information is placed in the parity field. The length of the parity field has yet to be determined

but is directly related to the length of the information field and the bit error rate required for the address field. The BER for the address and control fields should be at least two orders of magnitude better that the overall network BER of 10^{-7} in order to guard against misrouted or dropped packets. Thus, an overall BER performance of approximately 10^{-9} is required for the address and control fields. The information field will receive this link quality by default.

The idea of increasing the data content in the address and control fields and using two for three majority voting to guarantee 10^{-9} BER performance in those fields was contemplated, discarded and replace with the concept of using added parity. This was done to reduce the complexity of the onboard processing. It is assumed that the information data field will have to be encoded in order to maintain an overall end-to-end BER performance of 10^{-7} regardless of how the address and control fields are treated. Therefore, it appears to be more efficient to combine the address, control, and information fields together before encoding on the ground. Although a formal analysis has not been done, it appears that less parity bits are required to encode all three fields than to triple the address and control fields for use in majority voting and still require parity bits for encoding the information field -- albeit not as heavily as the combined encoding

requires. In addition, by heavily encoding the combining the address, control and information fields, no two-for-three majority voting circuits need be implemented.

TDM Frame Structure

The TDM frame structure has yet to be defined in detail. The TDM frame will be between one and 32 milliseconds in length. The frame efficiency increases with frame length. However, a longer frame requires greater onboard storage capability. Also, the packet length will directly effect the frame length. Since the downlink location capacity is limited by the size of the packet and the dwell time, if the packet size is large, the dwell time and frame length must be large in order to handle a reasonable number of packets per downlink dwell location. In addition, by making the dwell time and frame length as long as possible the hopping beam antenna system will not be required to switch as often, thus, improving system efficiency.

A superframe structure will be place over the TDM frame structure. Various orderwire messages will be reserved for particular frames within a superframe.

Downlink Orderwire Message Format

Orderwires will be used to convey satellite switch status, system timing information, initial access granted and denied messages, etcetera. The downlink orderwire message will be the first message of each dwell.

Contention and Congestion Control

In a destination directed packet satellite network, contention and congestion control are major concerns.

Contention problems appear in the NxN beam-to-beam switch. The beam-to-beam switch along with the MCDD-to-switch buffer must be design so that contention is avoided within this portion of the switching system (i.e. two or more inputs may not attempt to route to the same output at the same time).

Congestion occurs when more information is destined for a specific downlink/dwell than is available. This occurs because the data packets are self-routing and the routing information is not available until the packet arrives at the satellite. Because of the long propagation delay from the satellite to earth (125 msec), handshaking and requests for retransmission are impractical. In addition, since there is limited storage capability on the spacecraft, buffering of numerous packets for thousands of user is also impractical. Therefore, a congestion control method has to be developed that is specific to this destination directed packet switched satellite system.

Presently, two methods have been identified to deal with this problem.

The first method deals with this problem by simply denying access into the network based upon an analysis of the current state of the switching system and a statistical prediction of the additional capacity that would be required by the new user. In this scenario, during initial access, the user would inform the network control as to the destination of the message, the anticipated mean, mode, and peak data throughput requirements, and a request for point-to-point, multicast, or broadcast service. The network control would then take this information and determine whether or not there was enough capacity available to support the request. Using this method, the packets would be destination directed; however,

each packet would have to be sent to the same destination. Anytime the destination changes, a new request for access must be performed. There are two major drawbacks to this congestion control method: it would be extremely computational intensive to keep track of the statistical nature of each user's data, and the method precludes multiplexing users at the ground terminal.

A second method relies on distributive flow control at the ground terminals. In this scenario, the network control will continually monitor the downlink burst buffers to determine the current capacity of each downlink/dwell location. The network control will periodically transmit information regarding the current state of the downlink buffers to the ground terminals. This information will indicate the relative capacity of each downlink/dwell. For instance, downlink beam one dwell location three may be at 70 percent capacity while downlink beam seven dwell location six is at 90 percent capacity. The network control will set a threshold for capacity at perhaps 85 percent. Once that threshold is exceeded, no new transmissions are permitted to that downlink/dwell location. Communications that where already in progress to downlink beam seven, dwell six are allowed to continue; however, no new transmissions may be sent to this location until the capacity falls below 85 percent. Meanwhile, any user may transmit to downlink beam one dwell location three since its capacity is already under the 85 percent threshold. It is up to the ground terminals to institute the flow control. The threshold is set via the network control in order to allow the ground terminals adequate time to institute flow control before there is a congestion problem in the downlink burst buffer. One advantage that

this method has over the previous is that only the downlink burst buffers need to be monitored in order to determine the state of the switch instead of compiling statistics for every user in the system. A second advantage is that this method allows individual packets to be routed to different destinations; thus, enabling multiplexing of users at the ground terminal. One potential disadvantage may be that the threshold would have to be set to such a conservative number that the satellite capacity may be under severely utilized.

Network Hardware

Ground Terminals

The ground terminal is composed of and indoor and outdoor units (fig. 4).

The indoor unit consist of a terrestrial interface, protocol converter, packet assembler, encoder, continuous modulator, burst demodulator, decoder, message assembler, orderwire processor, and timing and control circuity.

The ground terminals will interface to the terrestrial telecommunications network at the DS0 (64 kbps), ISDN basic service rate, 2B+D (144 kbps), and T1-type rates (1.544 Mbps or 2.048 Mbps). In addition, the ground terminal will be capable of interfacing to commercial communications equipment and will be compatible with commercial standards.

The protocol converter provides an interface between the commercial communication packet switching standards and the internal packet switching protocol. All hand shaking, acknowledges, and flow control with the terrestrial networks will occur here.

The packet formatter breaks (or

appends) commercial packets into packets of constant length. The source and destination address and the control fields are appended to the fixed length packets and this total package is encoded. At this point, the synchronization bits are appended to the front of the packet and the information is passed on to the modulator.

The modulator and demodulator are two completely separate units. Presently, it is envisioned the uplink modulator will produce an offset QPSK signal and transmit continuously at either 64 kbps or 2.048 Mbps. Filtered OQPSK is used on the uplink in order to obtain a bandwidth efficiency of approximately 1.45 to 1.6 bits/sec/Hz. On the downlink, a burst demodulator is required. The modulation format has yet to be determined, but the data

rates will be in the 150 - 180 Mbps range.

The message assembler reads the demodulated data, strips off the source address fields and reassembles the orderwire messages and any messages destined for that ground terminal. The reassembled messages are then passed on to either the order wire processor or the protocol converter for entry into the terrestrial communications network.

Since this communications network utilizes time division multiplexing on the downlink with bursted data transmission in the 150 Mbps region, the timing and control of the communication is critical. The timing and control system (T&CS) is responsible for obtaining and maintaining synchronization with the satellite. The T&CS informs the
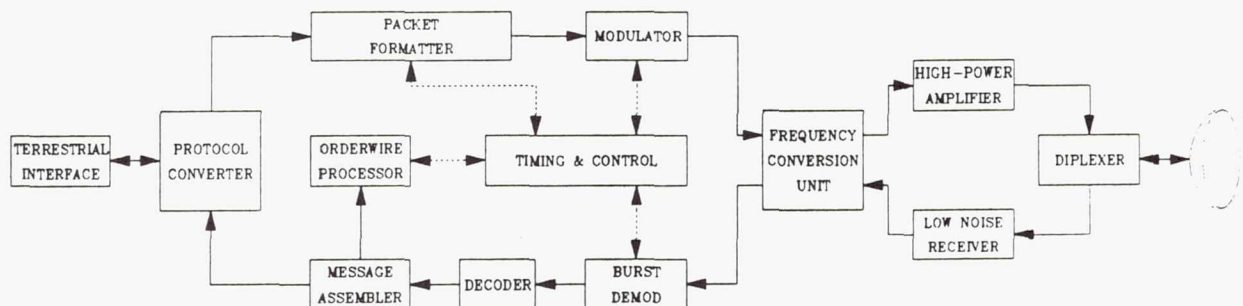
Figure 4 Ground Terminal

16

burst demodulator of the approximate time of burst arrival and receives a signal indicating actual burst arrival times. The T&CS uses the information obtained from the demodulator to adjust the ground terminals receive side timing in order to synchronize the ground terminal to the network. The T&CS also receives network control information from the orderwire processor and uses this information to determine what type of information is entered into the control fields of the transmitted packets. In addition, the T&CS will turn off the transmitter and modulator during periods where the ground terminal has relinquished access to the satellite uplink channel.

The outdoor unit contains the RF equipment consisting of the frequency conversion system, high power transmitter, diplexer, antenna system, and low noise receiver. The HPA is required to produce approximately 2 watts of transmit power. The required noise figure for the LRN is approximately 2.6 dB. The outdoor unit comprises the majority of the ground terminal cost.

MCDD

On-board demultiplexing and demodulation of narrowband traffic will be provided by multichannel demultiplexer demodulators (MCDD). In general, the MCDD can be viewed as a multifrequency channelizer and a demodulator system. The channelizer operates relatively independent of the modulation scheme; although some optimization for the channelizer may be performed if the modulation format has been identified early on. The demodulator system is either a time shared demodulator, a bank of individual demodulators, or a combination of the two.

The MCDD has been identified as a critical subsystem which needs to be developed for a FDMA/TDM architecture. Acousto-optical, optical, and digital signal processing technologies have all been identified as candidates for implementing a MCDD. NASA is investigating each of these approaches through contacts, grants, and in-house activity.
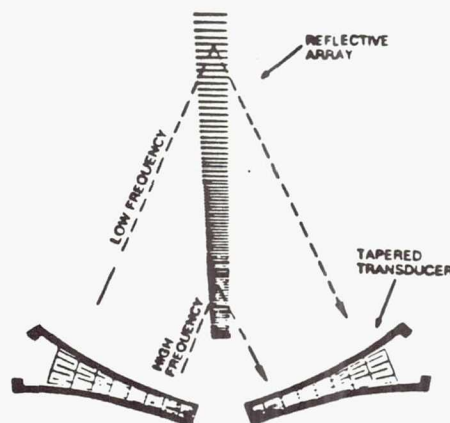
Figure 5 Hyperbolic Reflective Array Compressor

Amerasia Technology Incorporated is in the second phase of a Small Business Innovative Research Contract, NAS3-25862, to developing a proof-of-concept (POC) multichannel demultiplexer (MCD). The MCD uses a convolve-multiply-convolve technique to perform the demultiplexing function and is implemented using a surface acoustic wave herring-bone shaped reflective array compressor with hyperbolically shaped transducers (fig. 5).

Westinghouse Electric Corporation Communications Division is under contract to NASA LeRC (NAS3-25865) to develop a POC MCD which demonstrates the capability of demultiplexing 1000 low data rate FDMA uplinks. The multichannel demultiplexer is implemented as an coherent acousto-optic RF spectrum analyzer utilizing
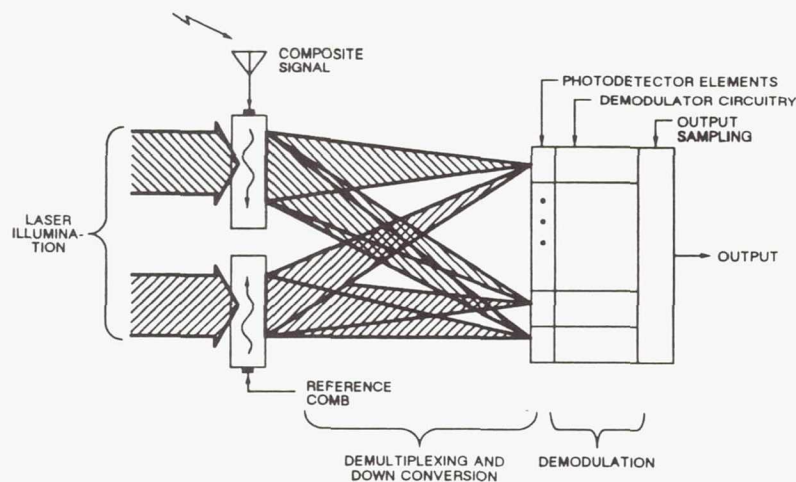
Figure 6 Optical MCDD

heterodyne detection with a modulated reference (fig 6). Similar to the RAC SAW implementation, the optical MCD is expected to have superior size, weight, and power requirement than a fully digital MCD and does not require a high speed A/D converter at the front end. The POC model will have a dynamic range of approximately 80 dB and will be capable of demultiplexing one thousand 64 kbps channels at 1.6 bps/Hz. Since the majority of the components are passive acousto-optic devices, this implementation of the MCD is highly reliable and radiation hard. Demodulation would be performed either serially, using a time shared demodulator, or in parallel, using an individual demodulator for each channel. One drawback to this implementation, however, is that a separate MCD is required for each separate data rate.

TRW is under contract with NASA LeRC (NAS3-25866) to develop a POC multichannel demultiplexer/ demodulator (MCDD) using advanced digital technologies. The composite FDM signal is A/D converted and channelized into wideband channels of 2.048 MHz bandwidth. The wideband channel is then either further channelized into 32 narrowband 64 kbps channels or passed directly on to the multirate demodulator as a 2.048 MHz channel. The modulation format used is differentially encoded OQPSK and the overall bandwidth efficiency of this system is 1.42 bps/Hz. The multirate demodulator can demodulate either one 2.084 MHz channel or thirty-two 64 kbps channels. This demodulator is designed as a continuous demodulator (fig. 7).

The University of Toledo is in the third year of grant (NAG3-799) to develop a programmable architecture for multicarrier demodulation based on parallel and pipeline digital design techniques for increased throughput. The hardware architecture and designs have been optimized for variable channel rates and variable numbers of channels. A POC model to demonstrate small-scale operation is under development.

LeRC has begun an in-house effort to develop and MCDD using commercial digital signal processors. The multichannel demultiplexer will be implemented as a combination of

18

software executing on general purpose DSP and state-of-the-art application specific DSP.

Demodulator

Once the demultiplexing function has been completed, the individual channels have to be demodulated. The present approach is to time share a bank of demodulators with each demodulator being capable of handling 24 to 32 channels.

The demodulators are presently designed to operate on continuous transmissions. which relates well to circuit switched operations. However, for packet switching, transmission is bursty. Therefore, either the demodulators will have to be capable of receiving burst transmissions or the ground terminals must transmit at regular intervals so that the continuous demodulators do not lose lock. For continuous demodulators, the ground terminals will have to send dummy packets. If the demodulators are capable of receiving burst transmissions, no dummy packets would be required. In addition, a TDM overlay could be placed on the FDMA uplinks whereby any uplink channel could be shared by multiple ground terminals.

Packet Synchronizing Buffer

The packet synchronizing buffer is responsible for receiving data from the MCDD and assembling and aligning the packets for use by the shared decoder. Assuming that the MCDD uses a time shared demodulator, the information from the MCDD will be presented to the packet synchronizer in a bit interleaved TDM format. Each bit in the TDM frame will correspond to a particular uplink frequency channel. The packet synchronizer will buffer each user data stream to a length of 2N-1, where N is the number of bits in a packet. The packet synchronizer will examine each user data buffer to determine the beginning of a packet and pass the individual packets -- minus the synchronizing header portion of the packet -- on to the shared decoder.

The memory requirements of this subsystem are quite large, (2N-1)*K*L where K is the number of channels in each MCDD and L is the number of MCDDs in the system. If packets from individual ground terminals could be sent to the satellite so that the packets reached the satellite synchronously, the memory requirements could be reduced by approximately 50 percent. The majority of this improvement is due to the fact that and additional (N-1) bits per channel is no longer required in order to be certain a full packet is captured. A second savings is achieved because there is
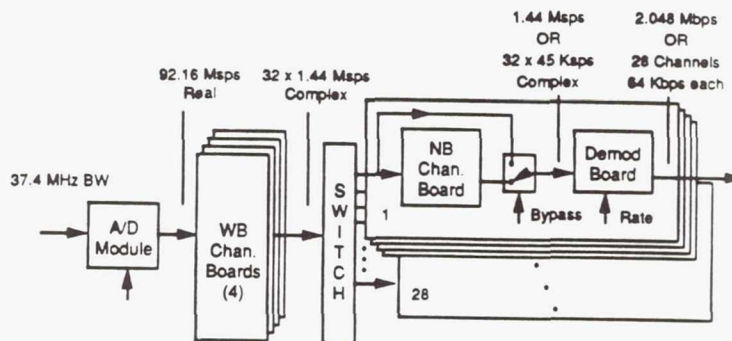
Figure 7 Digital MCDD

19

no longer a need for synchronization bits in the packet, thus reducing the overall packet length.

One possible method for synchronizing the uplink channels would be to the Global Positioning System (GPS). There are many commercial GPS receivers presently available. However, the cost is approximately $500 to $1,000 per ground terminal. By adding this additional complexity on the ground, the packet synchronizing buffer could be dramatically simplified.

Shared Decoder

The decoder subsystem decodes each packet on a packet by packet basis. Most likely, a bank of decoders will be time shared, particularly if the demodulator is time shared. Both, trellis and block decoders have been considered. If a trellis decoder were used, one could either throw away a predetermined amount of bits at the beginning of each packet in order to allow the decoder to initialize; or, one could save the previous state of the codec and jam this state into the decoder at the beginning of the next time slot for that particular source channel. For either of these methods, trellis decoding appears overly complex when consider the number or independent channels sharing the decoder. If a block decoder were used, the packet length would have to be an integer multiple of the block length. Since we have already determined that the packet will be a fixed length, the block code and packet length can readily be optimized.

Switching and Routing Elements

The switching and routing circuitry is composed of three major subsystems, the MCDD-to-Switch formatter, the 8x8 switch, and the Switch-to-TDM formatter. These three subsystem combine to effectively act as a 8192x64 packet switch and a 256x64 circuit switch assuming 8 MCDDs with either 1024 64 kbps users or 32 2.048 Mbps users. The MCDD-to-Switch and Switch-to-TDM formatters perform the temporal routing while the 8x8 switch performs the spacial routing.

MCDD-to-Switch Formatter
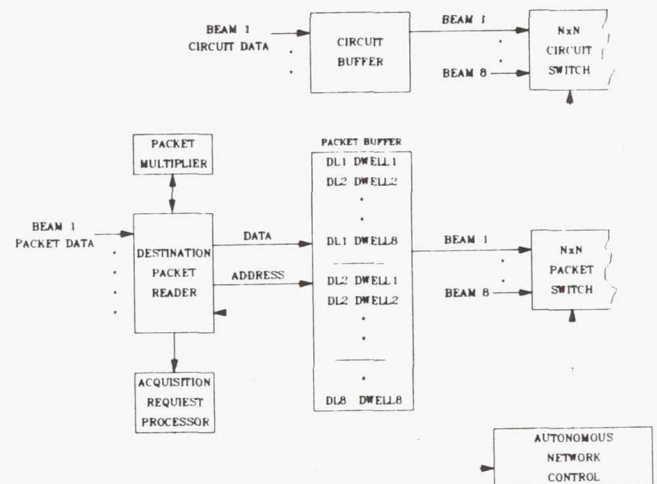
Figure 8 MCDD-to-Switch Formatter

The main function of the MCDD-to-Switch formatter is to take parallel messages and convert them into a TDM message stream (fig. 8). It receives decoded packets, examines the destination address, multiplies multicast and broadcast packets, sorts the messages, and stores the messages in a buffer for transmission through the NxN switch. In effect, the MCDD-to-Switch formatter acts as a 1024-to-64 switch with each message residing in the transmit buffer such that the messages can be transferred to the Switch-to-TDM circuitry in sequential order.

The circuitry for duplicating multicast and broadcast messages must reside in either the MCDD-to-Switch or the Switch-to-TDM formatters. Since the downlink beam address must be examined in the MCDD-to-Switch

20

formatter, it appears advantageous to put the packet duplication function here rather than in the Switch-to-TDM formatter. Therefore, the Switch-to-TDM formatter will only have to examine the dwell address and the downlink beam address can be discard.

NxN Switch

The NxN switch consist of two separate switches: the 2.048 Mbps circuit switch and the 64 kbps packet switch. This portion of the overall switching system is responsible for beam-to-beam interconnects. Since this is the sole function of the NxN switch, it is possible that both the circuit and packet data utilize the same type of switching fabric -- although not necessary. Regardless, both switches must be capable of handling contention problems relative to the downlink beams. Information from two separate inputs cannot reach the same output port at the same time. This problem must either be addressed within the NxN switch or be exclude from occurring by the MCDD-to-Switch formatter.

Of the overall switching and routing system, the NxN switch may be the most straight forward portion to implement. Numerous studies and papers have been published in this area (ref. 9-15). Optical switching and neural networks have also been recently investigate to solve this type of switching problem. One promising implementation is to use a high speed time-division-multiplexed fiberoptic bus to perform the NxN switching (ref. 16).

Switch-to-TDM Formatter

The Switch-to-TDM formatter must receive data from the eight ports of the spacial circuit switch and the spacial packet switch and write that information into the proper locations of the burst transmit memory (fig 9). This must occur without loosing any
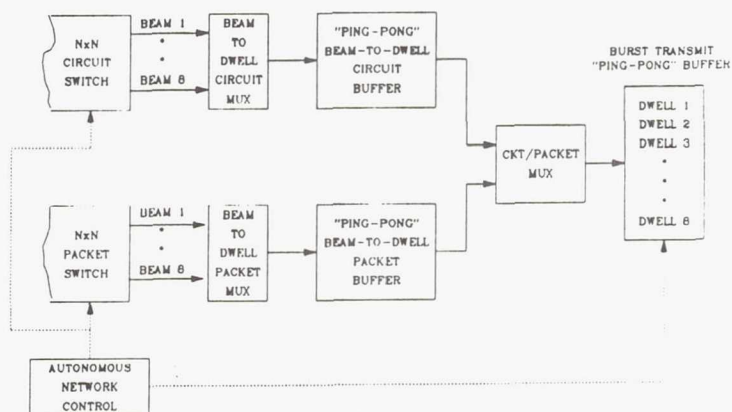
Figure 9 Switch-to-TDM Formatter

packets or circuits. Therefore, the Switch-to-TDM must be capable of resolving contention problems relative to the downlink dwell locations.

The burst transmit buffer is arranged so that each section corresponds to a particular downlink dwell location for the hopping beams. There must be reserved time slots within each dwell array for orderwires and for each 2.048 Mbps circuit destined for that particular downlink dwell. Additional memory space for each dwell is allocated by the autonomous network controller according to the "near-real-time" traffic demands of the packet network. After filling the appropriate dwell memory locations with circuit data, the Switch-to-TDM reads each packet and writes the packet to the corresponding memory location.

Encoder

The encoder is required to provide coding gain on the downlink. This encoder may be either a convolutional encoder or a block encoder capable of operating at 150 - 200 Mbps. A corresponding decoder is required at

the ground terminal. Presently, block decoders that handle these rates are available as commercial products. Convolutional decoders are much more difficult to implement. Therefore, the initial assumption is that a block encoder will be used.

Modulator

A burst modulator capable of a bandwidth efficient modulation scheme is required. A continuous phase modulation format is desired in order to run the satellite's high power amplifiers at saturation; thus, improving the downlink efficiency. NASA LeRC as an ongoing program in modulation and coding directed at such requirements. Among these are two completed contracts for 200 Mbps burst modems for satellite-to-ground applications: a 16 CPFSK modem, and an 8-PSK modem (ref. 17-18). Additional work is being performed by COMSAT Laboratories under contract NAS3-319317 for a programmable digital modem capable of binary, QPSK, 8-PSK, and 16 QAM modulation with up to 300 Mbps of data throughput.

Autonomous Network Controller (ANC)

On-board the satellite, the autonomous network controller (ANC) is responsible for allocation of the space and ground resources, and for real-time health monitoring and fault recovery of the on-board communication systems. The ANC may not perform all of the required network control functions; however, a favorable distribution of the network control functions will be realized between the on-board ANC and a ground-based network controller. In particular, traffic allocation and routing functions will be placed on-board to shorten call set-up and disconnect times. The ANC responds to narrowband user connection requests by allocating an uplink frequency to the requesting terminal.

The ANC will also allocates downlink time slots for 2.048 Mbps circuit switched data. The ANC will monitor the downlink burst buffers capacity, forward the burst buffer status to the ground terminals via downlink orderwires, and very the length of the downlink dwells to accommodate changing traffic patterns. In addition, the ANC will control the burst transmissions and the hopping beam antenna system.

CONCLUDING REMARKS

From an overall systems view, the problem of getting tens of thousands of low data rate users to communicate with each other through a processing satellite is of equal complexity whether it is accomplish using TDMA/TDM, FDMA/TDM, CDMA/TDM or another type of architecture. FDMA and more recently CDMA techniques have been touted as being superior to TDMA because of the reduced uplink transmit power required verses TDMA which, in turn, implies reduce ground terminal cost. These techniques, however, mandate that extremely complicated functions be performed onboard. In fact, all the functions from the MCDD to the transmit buffer of the MCDD-to-Switch formatter are necessary to get to a point that looks very similar to a TDMA uplink.

Any onboard processing system requires fault tolerant implementation. With size, weight, and power at a premium, traditional fault tolerant methods such as simple two-for-one redundancy of components and systems or majority voting will not suffice. NASA LeRC plans to address these issues in all aspects of the ISP design and is pursuing innovative fault-tolerant approaches which optimize redundancy requirements. Presently, the issue of fault tolerance in the digital multichannel demultiplexer is being address through a grant with the University of California, Davis

(NAG3-1166).

The present data rates of 64 kbps and 2.048 Mbps were chosen as a starting point and to be compatible with terrestrial ISDN networks. It is understood that these data rates may not be optimum. In particular, the uplink transmission rates will most certainly be sightly higher in order to accommodate the increase overhead inherent in a packet switched network.

STATUS AND FUTURE DIRECTIONS

NASA plans to develop a proof-of-concept (POC) information switching processor. The POC model will be constructed in-house at the NASA Lewis Research Center. In-house developed POC hardware will be supplemented by advanced fault tolerant components developed under contracts. The ISP architecture will ultimately be demonstrated in an satellite network simulation by integrating the ISP with high speed codecs, programmable digital modems, and multichannel demultiplexer demodulator currently being developed under industry contracts and university grants (ref. 19), and compatible ground terminals and on-board and ground based network control.

ACKNOWLEDGEMENTS

REFERENCES

1. Bauer, R.: ACTS High Burst Rate Link Evaluation Terminal, ACTS Quarterly, Issue 90/1, March 1990.
2. Ananasso, Fulvio and Saggese, Enrico: "User-Oriented Satellite Systems for the 1990's," AIAA 11th Communications Satellite Systems Conference Proceedings, 1986.
3. Price, K., Jorasch, R., and Wiskerchen, M.: Data Distribution Satellite, NASA CR-187066, Feb., 1991.
4. Bruno, R., Welti, G.: A Satellite System Concept for Fully-Meshed VSAT Networks, Stanford Telecommunications, April 1990.
5. Jorasch, R., Price, K. M.: Advanced Satellite System Architecture for VSATs with ISDN Compatibility, 12th AIAA International Communication Satellite Systems Conference, 1988.
6. Horstein, M., Hadinger, P.J.: A Satellite System Architecture for Single-Hop VSAT Networks, 12th AIAA International Communication Satellite Systems Conference, 1988.
7. Horstein, M., Hadinger, P.: Advanced Space Communications Architecture Study, NASA CR 179592, March 1987.
8. Roberts, L. G.: Dynamic Allocation of Satellite Capacity Through Packet Reservations, Proceedings of the National Computer Conference, NCC, 1973, pg. 711-716.
9. Berner, W. and Grassmann, W.: A Baseband Switch for Future Space Applications, Space Communications and Broadcasting, Vol. 5, pg. 71-78, 1987.
10. Ahmadi, H., and Denzel, W. E.: A Survey of Modern High-Performance Switching Techniques, IEEE Journal on Selected Areas in Communications, Vol. 7. No. 7, September 1989, pg. 1091-1103.
11. Wu, C., and Feng, T.: On a Class of Multistage Interconnection Networks, IEEE Transactions on Computers, Vol. C-29, No. 8. August 1980, pg. 674-702.
12. Uematsu, H., and Watanabe, R.: Architecture of a Packet Switch Based on Banyan Switching Network with Feedback Loops, Journal on Selected Areas in Communications, Vol. 6., No. 9, December 1988, pg. 1521-1527.
13. Spanke, R. A.: Architectures for Guided-Wave Optical Space Switching Systems, IEEE Communications Magazine, Vol. 25, No. 5, pg. 42-48, May 1987.
14. Hinton, H. S.: Photonic Switching Fabrics, IEEE Communications Magazine, April 1990, pg. 71-89.
15. Brown, T. X.: Neural Networks for Switching, IEEE Communications Magazine, November 1989, pg. 72-81.
16. Inukai, T., and Faris, F.: Advanced Satellite System Concepts Task $: Onboard Processing for VSAT Services, NASA Contract No. NAS3-25092, Interim Presentation, pg. 5-33,36.
17. Davis, R. C., Wernlund, J. V., Gann, J. A., Roesch, J. F., Wright, T., Crowley, R. D.: Advanced Modulation Technology Development for Earth Station Demodulator Applications, NASA Contractor Report 185126, July 1989.
18. Advanced Modulation Technology Development

for Earth Station Demodulator Applications, NASA Contractor Report CR18549, 1989.
19. Harrold, J.L., Budinger, J.M., Stevens, G.H.: "Onboard Switching and Processing," Proceedings of the IEEE, July 1990, pages 1206-1213.

A CODE PHASE DIVISION MULTIPLE ACCESS (CPDMA) TECHNIQUE FOR VSAT SATELLITE COMMUNICATIONS

R. Bruno, R. McOmber, A. Weinberg
Stanford Telecommunications, Inc.
Reston, Virginia

Abstract

This paper describes a reference concept and implementation relevant to the application of Code Phase Division Multiple Access (CPDMA) to a high capacity satellite communication system providing 16 Kbps single hop channels between Very Small Aperture Terminals (VSATs). The description includes a potential implementation of an on-board CPDMA bulk demodulator/converter utilizing programmable CCD technology projected to be available in the early 90's. Also provided are a high level description of the system architecture and operations, identification of key functional and performance requirements of the system elements, and analysis results of end-to-end system performance relative to key figures of merit such as spectral efficiency.

I. Introduction

In recent years, a great deal of effort has gone into developing advanced communications satellite architectures and concepts that are suited to providing single hop service between Very Small Aperture Terminals (VSATs) at user premise locations. Satellite transponders work most efficiently in a Time Division Multiple Access (TDMA) mode whereby a satellite transponder is time-shared by a number of users and each user is served via dedicated short duration bursts at high data rates. From a consideration of the satellite alone, TDMA allocation among the user population is the most efficient operating mode for both uplink and downlink channels. However, high burst rate TDMA on the uplink is incompatible with VSAT cost constraints. In recognition of the need for low cost VSATs, many studies of advanced satellite communications systems serving VSATs have converged on a concept of Frequency Division Multiple Access (FDMA) Single Channel Per Carrier (SCPC) on the uplinks and burst TDMA on the downlinks. The FDMA/SCPC scheme on the uplink supports the concept of a low cost VSAT, but the FDMA/TDMA conversion required at the satellite represents a significant processing burden. The Code Phase Division Multiple Access (CPDMA) concept was developed as an alternative to FDMA/TDMA to support future commercial communications from VSATs. The CPDMA techniques for separation of distinct uplink channels from VSATs meshes particularly well with TDMA downlinks in that the required on-board CPDMA/TDMA demodulation and conversion is easily accomplished with little computational complexity.

II. The CPDMA Concept

Figure 1 conceptually illustrates the CPDMA data baud structure at baseband. The N VSAT users transmit continuously using an identical maximal length PN code of length L and fixed PN chip duration T_c. The user symbol duration is fixed at one period of the PN code:

$$T_s = LT_c \tag{1}$$

At the satellite front-end, all symbol epochs are aligned in time and the required signal bandwidth is approximately $2/T_c$. The individual VSAT users are distinguished by distinct PN code epochs (equivalently, code phases). In the figure, a code phase separation between adjacent VSAT users of one chip is illustrated but, more generally, the code phase separation can be $\geq$ one chip.

Figure 2 is a high level diagram of a satellite CPDMA demodulator implementation based on binary DPSK modulation. The demodulator begins with generation of baseband in-phase and quadrature signal waveforms via downconversion by a non-coherent local oscillator. These baseband waveforms are then input into chip matched filters whose outputs are sampled and fed into PN Matched Filters (PNMFs). A simple hardware implementation of the PNMFs using CCDs is described in Section 3 below.

The role of each PNMF is to sequentially correlate the input chip samples over one symbol period with each code phase of the original PN code sequence. At each code phase position, the PNMF reference will match one and only one user

REFERENCE:

SYMBOL #1 ———————→|←——————— SYMBOL #2

P_1 P_2 P_3 • • • • • • P_{L-1} P_L | P_1 P_2 P_3 • • • • • • P_{L-1} P_L | • • •

USER 1: P_1 P_2 P_3 • • • • • • P_{L-1} P_L | P_1 P_2 P_3 • • • • • • P_{L-1} P_L | • • •

USER 2: P_L P_1 P_2 • • • • • • P_{L-2} P_{L-1} | P_L P_1 P_2 • • • • • • P_{L-2} P_{L-1} | • • •

USER 3: P_{L-1} P_L P_1 • • • • • • P_{L-3} P_{L-2} | P_{L-1} P_L P_1 • • • • • • P_{L-3} P_{L-2} | • • •

USER L: P_2 P_3 P_4 • • • • • • P_L P_1 | P_2 P_3 P_4 • • • • • • P_L P_1 | • • •
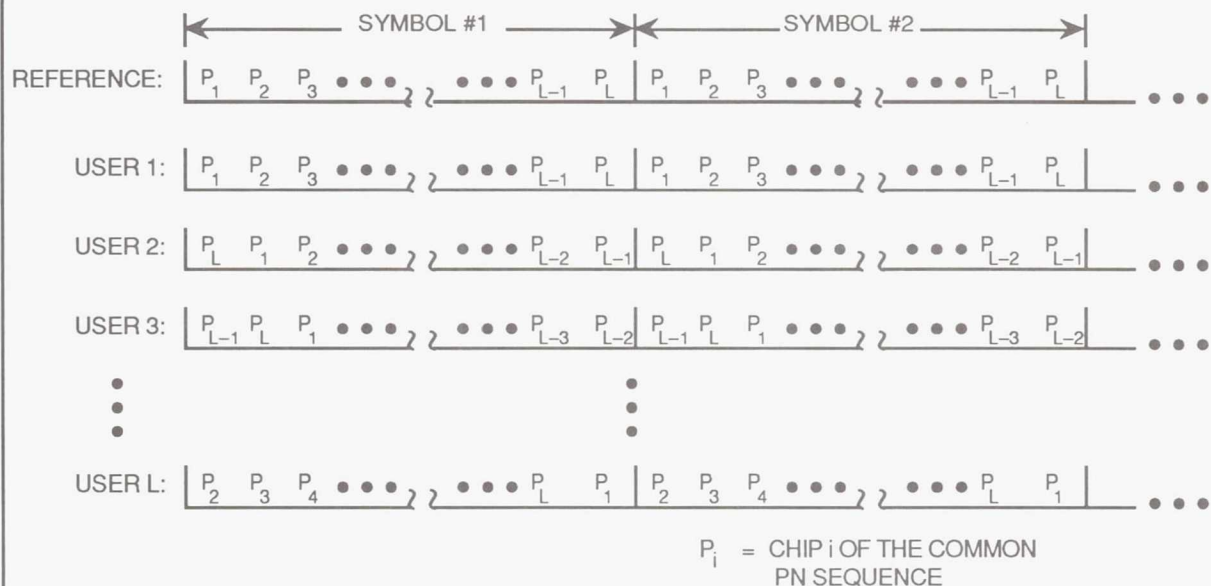
P_i = CHIP i OF THE COMMON
PN SEQUENCE

FIGURE 1 CPDMA DATA BAUD STRUCTURE FOR A USER CODE
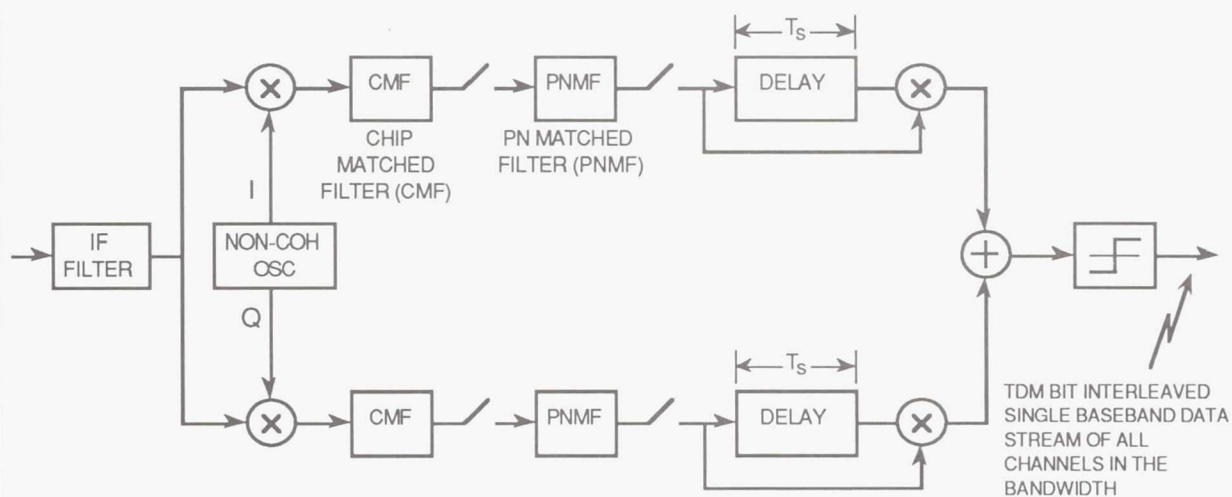PHASE SEPARATION OF ONE PN CHIP

FIGURE 2. CHARGE COUPLED DEVICE (CCD)
DEMODULATOR/CONVERTER — BINARY DPSK

26

code phase with all other users producing minimum correlation. The sequence of correlations within the PNMF thus produce a sequence of despread samples corresponding to each user's data symbol in turn. These samples are then processed using conventional DBPSK methods to yield a single Time Division Multiplexed (TDM) baseband data stream consisting of all symbols from all users.

Because the user symbol boundaries are aligned in time and each symbol consists of one full period of the code sequence, the correlation properties of the Maximal Length PN sequences are not disrupted by the presence of user data. Under ideal conditions (e.g., no user signal time or frequency errors) the cross correlation between one user signal and any other user signal is $1/L$ in magnitude. With a total of N users, each with signal energy E, there will be N-1 users interfering with the desired signal, producing a composite interference signal at the PNMF output having zero mean and variance KE/L^2. Treating the composite interference signal as additive gaussian noise, an effective E/N_o ratio due to user-to-user interference can be computed for the CPDMA system:

$$\text{CPDMA: } (E/N_o)_{eff} = \frac{E/N_o}{1 + \left(\frac{N-1}{L^2} \right)(E/N_o)} \quad (2)$$

Compare this result to conventional Code Division Multiple Access (CDMA) in which user PN codes are not synchronized with transmitted symbols. In this case, the composite of the N-1 other users supplies an interference energy of $(N - 1)E$ over the user bandwidth for an effective E/N_o of:

$$\text{CDMA: } (E/N_o)_{eff} = \frac{E/N_o}{1 + \left(\frac{N-1}{L} \right)(E/N_o)} \quad (3)$$

Equation 2 implies that, for a code length L sufficiently large, the CPDMA system can support up to L users spaced one chip apart in code phase without significant degradation to effective E/N_o. Under ideal conditions, CPDMA represents a significant improvement over conventional CDMA.

The above results assume ideal conditions of no user time and frequency errors relative to the satellite references. In a practical system, the VSAT users will not have perfect knowledge of time and frequency, and time errors may be significant relative to the duration of a code chip. For this reason, it is desirable to separate VSAT users by more than one code chip in phase — at the expense of fewer users per CPDMA group and reduced bandwidth efficiency. To date, our studies have focused on a user code phase separation of 1.5 chips.

Section 3 describes a CPDMA bulk demodulator implementation using Charge Coupled Devices (CCDs). While the description is based on a 1.5 chip user separation, the fundamental architecture is applicable to other code phase separations. Section 4 describes system performance for the baseline 1.5 chip separation while taking into account less than ideal user conditions.

III. CPDMA Bulk Demodulator Implementation

As described above, the PNMF must sequentially correlate the received composite signal samples output from the chip matched filter with each phase of the code sequence and output a time division multiplexed stream of samples representing the stream of I or Q symbols for all users. A potential CPDMA implementation of the PNMF based on CCD shift registers has been defined and is illustrated in Figure 3 for a user code phase separation of 1.5 chips. Input to the PNMF are analog samples of the chip matched filter (CMF) output waveform which contains a composite of all user signals. The sample rate is twice the PN chip rate, so a CCD register of length 4L holds two symbols worth of samples. Note, however, only one of the symbols will be processed, so two such PNMFs on each of the I and Q channels are required. Every other stage of the signal register is tapped — resulting in 2L total taps. The second shift register holds two complete sequences of the PN code which remain fixed as the user signal samples pass through their register. Finally, a third shift register contains a block of L one's which are shifted right one position for every two CCD sample times.

Because the PN code register is fixed, movement of the CMF output samples through the CCD register will cause multiplication of the signal samples by a different phase of the PN code each sample time. Since the user code phases are separated by 1.5 chips (or three samples), the stored tap weights will match a different user's PN code phase every third CCD sample.

The role of the mask word is to limit accumulation to samples within a single symbol interval. The

27

FIGURE 3. CANDIDATE CCD IMPLEMENTATION OF PN MATCHED FILTER FOR USER CODE PHASE SEPARATION OF 1.5 CHIPS
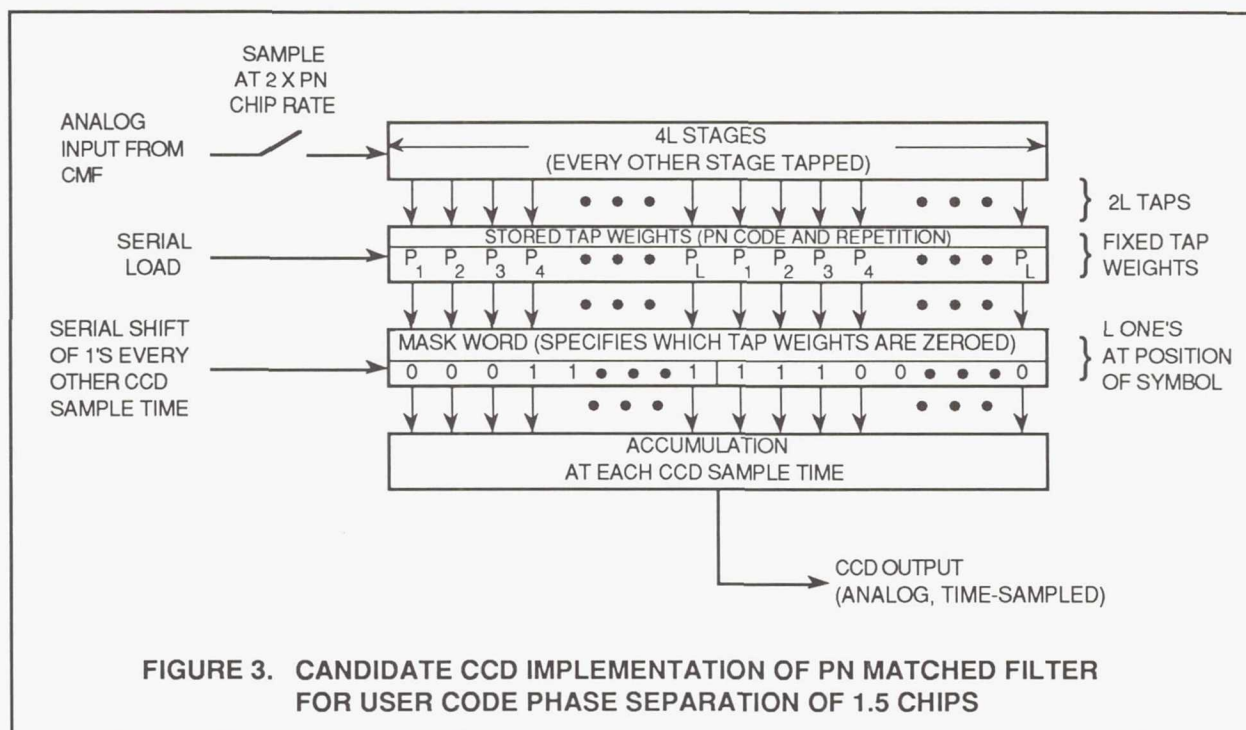
mask word (consisting of L one's) moves down its shift register at a location and rate which match the symbol interval within the signal sample register. As time progresses, the non-zero samples input into the accumulator will correspond to one symbol's worth of the composite samples output from the CMF multiplied by each phase of the PN code in turn. The output from the accumulator is a correlation over one symbol time of the received composite waveform of all user signals with each phase of the PN code.

As stated above, the stored tap weights match a different user's PN code phase every third signal sample. The accumulator output sequence contains the on-time I or Q channel symbol sample for a different user every third sample. However, the intermediate samples are not wasted. At each sample time, the relative phases between the received signal samples and reference PN code shift by one-half chip. The intermediate samples represent correlations of each user's symbol with a reference either one-half chip early or one-half chip late from the ideal user symbol/PN epoch required. The satellite uses these samples to derive an uplink timing error measurement for each VSAT user which is then sent to the users via the downlink. It is the user's responsibility to use this information to maintain required uplink timing accuracy.

For a user code phase separation of 1.5 PN chips and a code length L, the number of CPDMA user channels which can be processed is $L/1.5$ channels per demodulator. With sampling at twice the PN chip rate, the CCD sample rate required is $2L/T_c$ samples per second. For example, in a rate 1/2 encoded DBPSK implementation providing user services at 16 Kbps, a length 255 PN code could be used to permit simultaneous demodulation of 170 CPDMA user channels with a CCD sample rate of 16.32 MHz. If the code length is increased to 1023 chips, 682 channels are provided but the CCD sample rate required increases to ~65 MHz. Technology limitations on achievable CCD sample rate are one factor limiting the maximum number of user channels which can be demodulated by a single CPDMA bulk demodulator.

IV. A CPDMA Reference Architecture

Figure 4 illustrates one potential implementation of a satellite providing connectivity among a large number of VSAT users. On the uplink, eight fixed antenna beams are used to receive VSAT transmissions. Each antenna beam is divided into M CPDMA user groups with individual groups having distinct carrier frequencies and PN codes. On the downlink, TDMA is used on four hopping beams each having a single downlink carrier.

28

FIGURE 4. SAMPLE CPDMA/SATELLITE SWITCHED/TDMA APPLICATION
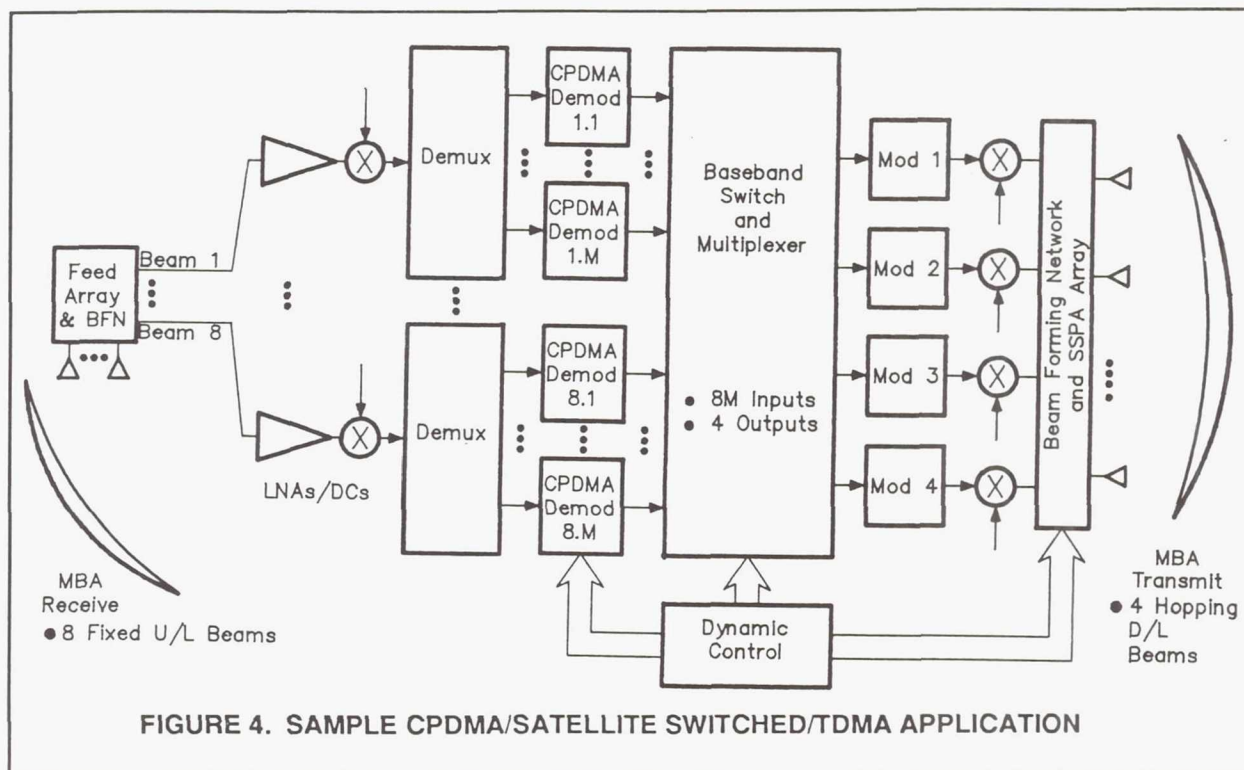
Continuing the rate 1/2 encoded DBPSK example from above, for a code length of 255 chips, there are 170 users per CPDMA group and, if M=33 bulk demodulators per beam are used, there will be 33 * 170 or 5610 user channels per beam or 44,880 total user channels. At a data rate of 16 Kbps, the total throughput available is ~718 Mbps and each downlink beam must operate at ~180 Mbps. If the PN code length is increased to L=1023 chips, then only 8 to 9 CPDMA bulk demodulators per beam are required to provide this level of service.

In addition to demodulation of all VSAT transmissions, the satellite also provides data buffering and routing to the appropriate downlink beam, relay of channel control and channel status messages between the VSAT and network controller, and support for the uplink acquisition process executed by the VSAT users. The uplink acquisition process is described in some detail below.

VSAT responsibilities include acquisition and tracking of the satellite provided TDMA downlink, uplink acquisition and maintenance, generation and reception of service control messages to and from the network controller, and communication service implementation and monitoring. Because accurate uplink timing is essential for operation of the CPDMA system, the VSAT must reduce local time

errors by acquiring and tracking the satellite downlink before attempting uplink acquisition. Following acquisition, the VSAT must still maintain accurate uplink timing by compensating for satellite motion using satellite ephemerides provided by the network controller, and by adjusting uplink timing in response to uplink tracking measurement reports provided by the satellite.

Even with tracking of the satellite downlink, VSAT uplink timing errors are expected to be too large to permit direct access of a specified CPDMA channel within the group used for VSAT-to-VSAT communications. Instead, an uplink acquisition procedure must be executed by the VSAT using a separate set of CPDMA channels provided by the satellite. The acquisition channel group needs to have enough CPDMA phase channels both to cover the uplink time uncertainty (or equivalently PN phase uncertainty) region of the VSAT uplink signal and to make contention between VSAT users attempting uplink acquisition unlikely.

The acquisition process starts with the VSAT sending uplink probes into the acquisition channel group. These probes can be modulated with information and need to contain at least enough information to identify the VSAT user attempting acquisition. Assuming no contention with another VSAT user at the same code phase, the satellite

29

demodulates the received acquisition probe to identify the user and, additionally, measures the PN phase of the probe. This phase information is sent to the VSAT in order to allow the VSAT to make appropriate uplink clock adjustments. The probe/response cycle may need to be repeated several times before the VSAT clock uncertainty is reduced sufficiently to permit assignment of a communications channel.

A network controller is also required in the reference architecture. The network controller is responsible for assignment and control of all satellite communications resources as well as monitoring of both satellite performance and service operations. Additionally, the network controller must perform satellite tracking and generate satellite ephemerides for distribution to the VSAT users.

V. CPDMA Performance Issues/Evaluations

Initial CPDMA uplink performance examinations have been conducted for the reference architecture using a code length of L=255 chips. Of particular concern were the overall bandwidth efficiency obtainable in the CPDMA architecture as well as oscillator stability and tracking accuracy requirements imposed on the VSATs by the implementation.

Noise sources impacting the VSAT transmitted signal include: Additive white gaussian noise, interference from other user signals within the user's CPDMA group, and interference from users in other CPDMA user groups. The signal transmitted by a VSAT user is subject to time and frequency errors which reduce the signal correlations detected at the satellite. Additionally, channel filtering used to limit interference between CPDMA user groups also reduces the detected signal level. Note, however, that these signal degrading effects are not unique to a CPDMA implementation, but are similar for an ordinary PN spread communications link. For this reason, the current discussion is limited to the interference sources only (although bandwidth efficiency results presented below include all degrading factors).

Of greatest interest is the interference contribution from users within the desired user's CPDMA group. Users within a common CPDMA group are separated only by PN code phase. VSAT uplink signal time errors must be kept small to prevent significant cross-correlation between users that are adjacent in phase. Additionally, factors

which could disrupt the ideal correlation properties of the maximal length PN sequences used must also be considered.

Our analyses have indicated that the interference between users adjacent in code phase is small as long as user uplink timing errors are held below 10 to 20 percent of a chip duration. While channel filtering to limit required bandwidth does increase interference among users within a common CPDMA group, the 1.5 chip phase separation between adjacent users keeps the impact small. For the reference architecture, with a length 255 PN code and 16 Kbps data rate, the 0.1 to 0.2 chip accuracy requirement translates into an absolute time error of 12 to 25 ns. To maintain this timing the VSATs must use the timing error measurements periodically reported on the satellite downlink.

Under ideal conditions of zero time and frequency error, the correlation between a user signal and any other user's signal within a CPDMA group is 1/L. For users separated significantly in PN code phase from a desired user, time errors do not alter this result. However, uplink signal carrier frequency errors can significantly raise the correlation magnitude. Our analyses have indicated that maintenance of low signal correlations among users within a common CPDMA group is the driving factor behind user requirements on carrier frequency stability. To maintain the correlation minimums of the maximal length PN sequence the user signal frequency error (Δf) relative to the satellite reference needs to be kept below:

$$\frac{\Delta f}{f_c} < 10^{-8} \qquad (4)$$

where

f_c = Nominal Carrier Frequency

To maintain this carrier frequency accuracy the VSAT must track the frequency of the satellite downlink and compensate for satellite motion using ephemeris predictions. Additionally, the satellite will need to maintain a stable local reference. This reference could be obtained by tracking a stable reference provided by the network control station.

Also included in our evaluations were the degradations due to interference from users in other CPDMA groups. Users in other CPDMA groups are separated from the desired user by both PN code and frequency. It is assumed that users

would provide uplink channel filtering to limit the spectral occupancy of their transmissions. Of key importance is the frequency spacing required between CPDMA groups since this spacing determines the overall bandwidth efficiency achievable for the reference architecture.

The current study has been limited to examination of channel spacing and channel filter trades for Butterworth type filters only. Using computations of uplink bit error rate which included all of the degrading factors mentioned thus far, we explored performance as transmission filter bandwidth and user group frequency were varied. The results of this analysis indicate that a center-to-center frequency spacing equal to 0.75 times the null-to-null PN spread signal bandwidth of $2/T_c$ Hz can be readily achieved. Based on this frequency separation between CPDMA channel groups, Table 1 summarizes bandwidth efficiencies for the CPDMA reference architecture under a variety of modulation and coding options. We believe that the current CPDMA implementation, having a code phase separation of 1.5 chips between CPDMA users, provides a bandwidth efficiency that is competitive with FDMA bulk demodulator implementations. Also note that no attempt to optimize all of the parameters of the reference architecture has yet been performed. In particular, it may be possible to significantly reduce the current 1.5 chip user phase separation without sacrificing performance. A reduced phase separation permits more users within a CPDMA user group and raises bandwidth efficiency.

VI. Conclusions

A reference concept and implementation of a communications architecture using code phase division multiple access for single hop satellite communications between VSAT terminals has been described. The architecture includes use of bulk demodulators based on CCD technology projected to be available in the early 90's. Additionally, performance evaluations have been conducted in order to define key VSAT and satellite functional requirements as well as achievable spectral efficiency.

Table 1. CPDMA Achievable Spectral Efficiency

Modulation Format	Coding Rate	Spectral Efficiency (bps/Hz)
QPSK	Uncoded	8/9
QPSK	Rate 3/4	2/3
BPSK	Uncoded	4/9
BPSK	Rate 3/4	1/3

Page intentionally left blank

Mobile Telephony Through LEO Satellites: To OBP or Not

Paul A. Monte and Ming Louie
Space Systems/Loral
Palo Alto, CA

R. Wiedeman
Loral Aerospace Corporation
San Jose, CA

GLOBALSTAR is a satellite-based mobile communications system that is interoperable with the current and future Public Land Mobile Network (PLMN) and Public Switched Telephone Network (PSTN). The selection of the transponder type, bent-pipe or on-board processing (OBP), for GLOBAL-STAR is based on many criteria, each of which is essential to the commercial and technological feasibility of GLOBALSTAR. This paper describes the trade study that was done to determine the pros and cons of a bent-pipe transponder or an on-board processing transponder.

The design of GLOBALSTAR's telecommunications system is a multi-variable cost optimization between the cost and complexity of the individual satellites, the number of satellites required to provide coverage to the service areas, the cost of launching the satellites into their selected orbits, the ground segment cost, user equipment cost, satellite voice channel capacity and other issues. This paper focuses on the cost and complexity of the individual satellites, specifically the transponder type and the impact of the transponder type on satellite and ground segment cost, satellite power and weight, and satellite voice channel capacity.

Introduction to GLOBALSTAR

GLOBALSTAR is a satellite system which offers global mobile voice and data services and radio-determination satellite services (RDSS) to and from hand-held and vehicle-mounted transmit and receive devices. By combining the use of low-earth orbit (LEO) satellites with existing terrestrial communications systems and innovative, highly efficient spread spectrum techniques, the GLOBALSTAR system provides users throughout the world with low-cost, reliable communications. The system uses a constellation of 48 operating LEO satellites to provide optimum global coverage.

Because 90 percent of all traffic from a given point will be accommodated by a single gateway, GLOBALSTAR has been configured to link the mobile unit to a terrestrial gateway through a single satellite so that the system requires no satellite traffic crosslinks. GLOBALSTAR incorporates existing terrestrial communications facilities into its overall configuration through gateway earth station interfaces. The interoperability of GLOBALSTAR with the PSTN enhances the system's reliability and decreases costs to the end user by decreasing the complexity of the space segment. By complementing rather than supplanting existing carriers' networks and by sharing revenues with existing carriers, GLOBALSTAR can achieve rapid adoption throughout the United States and the world.

33

GLOBALSTAR proposes three alternative spectrum plans. For brevity, this paper will use only the one employing L-band with C-band feeder links. This system makes bidirectional use of the allocated RDSS spectrum in the L-band (1610-1626.5 MHz).

The GLOBALSTAR system is designed to operate compatibly with other LEO satellite systems providing RDSS, voice and data services and can operate without causing harmful interference to geostationary, RDSS systems, radio navigation systems and GLONASS. Moreover, the capacity of the GLOBALSTAR system will be only slightly degraded by operations from those systems.

Criteria & System Definition

The selection of transponder type, bent-pipe or OBP, was based on the simultaneous solution of many criteria: the complexity and cost of the individual satellites, the weight of the communications payload, the power requirements of the communications payload, the availability of equipment (or the amount of development required), satellite voice channel capacity, security (both for privacy and for fraud), and quality of service. There is also a trade-off in cost and complexity between the satellite and the gateways.

The 48-satellite constellation is the Walker 48/8/1 constellation (ref. 1) at an altitude of 1389 km. (750 nm.) with an inclination of 52 degrees. This constellation has 48 satellites in eight orbit planes, all with an inclination of 52 degrees. The phasing of the satellites from one plane to another is shifted by 7.5 degrees. This 48 satellite constellation provides 100% single coverage from the 65 degrees south latitude to 65 degrees north latitude with a minimum elevation angle (the angle from the horizon to the line of sight between the user and the satellite) of 10 degrees and provides 100% double coverage of the continental United States with a minimum elevation angle of 10 degrees with one satellite at a higher elevation angle than 15 degrees.

With the satellite constellation in a non-polar, inclined orbit, the coverage is constantly changing for a particular point on the earth (although it is predictable): sometimes the covering satellite is moving northeast, sometimes southeast; sometimes there are three or four satellites in view. As the coverage areas for the satellites move across one another, there will be self-interference (or the system has to be tightly controlled to prevent it). This will be taken into consideration in determining the transponder. There might also be interference from other operators.

With the satellite constellation defined, the satellite lifetime defined (7.5 years), input from market studies and input regarding the usage of cellular phones, a program was run to see the maximum number of circuits required during the busy hours in the seventh year of operation. In the seventh year, there will be approximately 1.5 million users which will require a satellite capacity of 1900 channels during the busy hour.

Code Division Multiple Access (CDMA) is the access of choice. CDMA achieves its high capacity by achieving more effective frequency reuse than other methods. This technique allows a statistical averaging principle known as the "Law of Large Numbers" to come into effect with the result that frequency reuse is more efficient than with Frequency Division Multiple Access (FDMA) or Time Division Multiple Access (TDMA) techniques. The CDMA technology used in this system exploits the following techniques in obtaining high spectral efficiency: voice activity, error detection and correction, efficient demodulation, antenna directivity (spot beams) and multiple satellites.

Spot beams are required to deliver the capacity required to and from the mobiles. GLOBALSTAR uses six elliptically shaped spot beams. The major axis of the elliptical beams are aligned with the velocity vector of the satellite movement in order to decrease the number of inter-beam handoffs. Figure 1 shows the six spot beams illuminating the United States.

The spot beams are designed to compensate for the difference in the satellite-to-user link losses between the "near" and "far" users, so that the power flux density of the "far" users is about the same as the "near" users (*isoflux* design). This antenna design reduces the near-far problem, decreases the range of power control required for CDMA and increases the capacity of the system. Figure 2 shows a cross cut of the antenna pattern.

Figure 1 **Spot Beams on the US**

Figure 2 **Isoflux Antenna Pattern**

The multiple access technique used for the system is Time Domain Duplex-Frequency Division-Code Division Multiple Access (TDD-FD-CDMA). The most efficient method of using L-band in both the user-to-satellite and satellite-to-user directions is TDD. The 16.5 MHz at L-band is divided into 13 sub-bands of 1.25 MHz CDMA channels with each channel being TDD. There are multiple CDMA users in a 1.25 MHz channel. The frequency plan of how the user uplinks are translated to the gateway downlinks is shown in Figure 3. Beam hopping is used to minimize the interference from one beam to another beam and decrease the interference to/from other systems. The system has a 60 msec TDD frame with six 10 msec time slots. Three time slots are allocated for transmit and three time slots are allocated for receive. Within an individual time slot, the signals will either transmit or receive two of the six beams (e.g. beams one and four, or beams two and five, or beams three and six). This is shown in Figure 4.

The system trade-off left the following limits on the satellites: the cost could not be more than $10 million per satellite; the DC power required by the communications payload could not be more than 750 watts; the satellite could not weight more than 500 kilograms.

A block diagram was designed for the transponders (shown in Figure 5) which shows where the OBP equipment would be placed.

Link Budgets

Link budgets were produced for the bent-pipe satellite and the OBP satellite to compare the capacity and power consumption of the two different transponders. The bent-pipe link budget is shown in Figure 6 and the OBP link budget is shown in Figure 7.

Each Figure has both the forward path (Gateway to User - columns B and H) and the return path (User to Gateway - columns C and G) as both paths need to be examined to see the trade-offs in the system regarding the link budgets. Lines six through 15 calculate the EIRP. In the ground to space links,
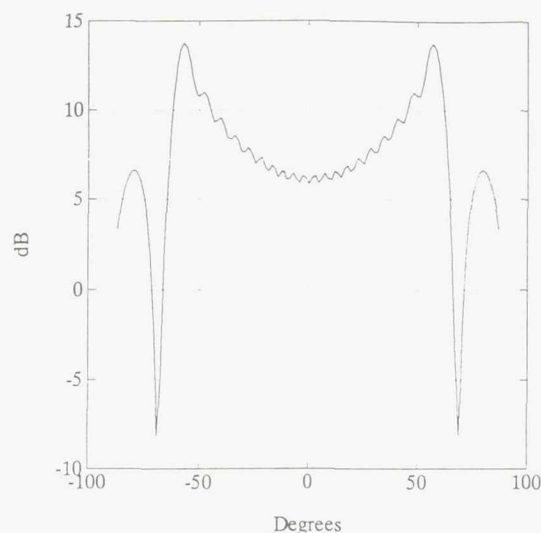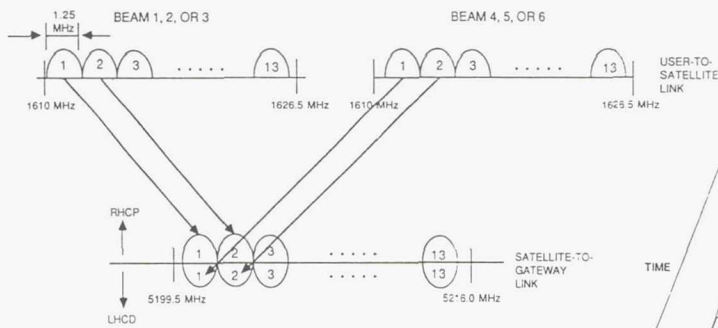
Figure 3 **Frequency Plan**
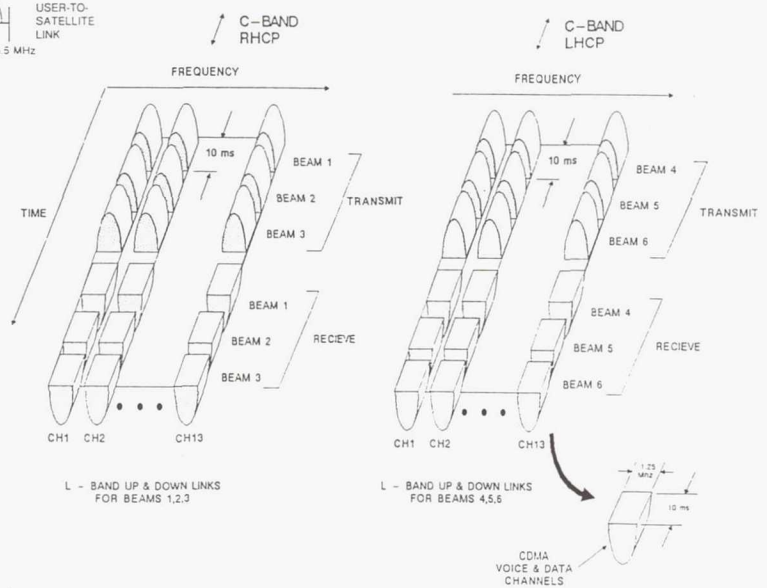
Figure 4 **TDD-FD-CDMA**
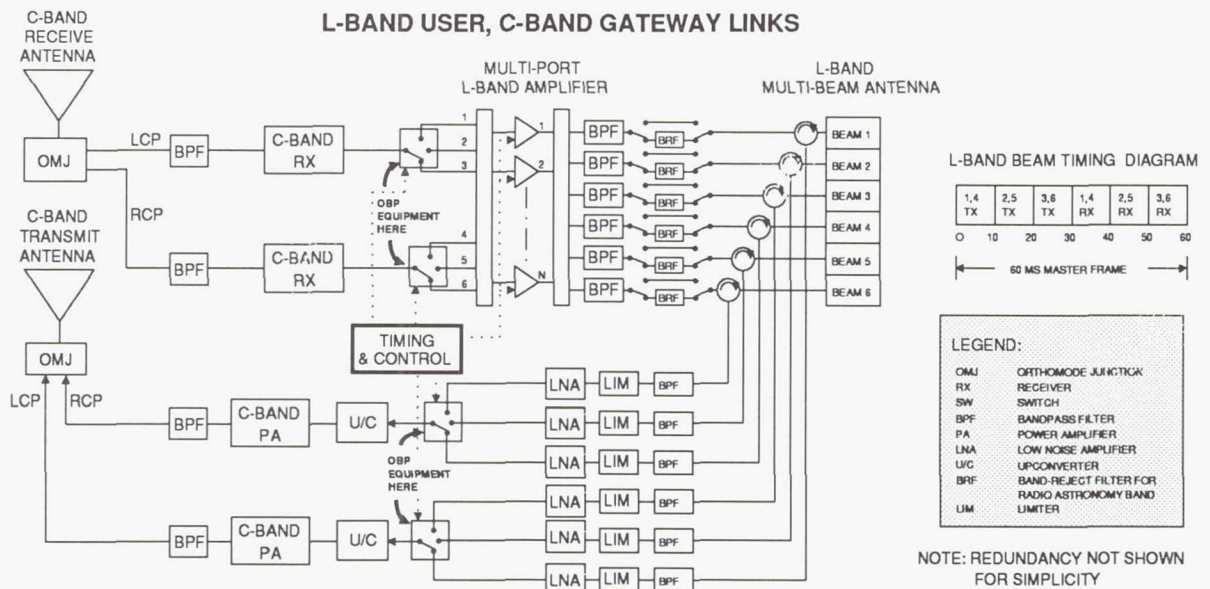
Figure 5 **Transponder Block Diagram**

the power is EIRP/user whereas in the space to ground links the power is EIRP/1.25 MHz channel. The number of users denoted is the number of users in a 1.25 MHz channel. Due to the beam hopping and TDD operation, a transmitter is only on for 1/6 of the time, therefore the transmitter power shown is the power required when the system is transmitting - not the average power.

As spacecraft antenna are isoflux, the elevation angle does not affect antenna gain minus space loss. The antenna gain increases as the space loss increases. The antenna gain shown is the gain with regard to the space loss at nadir (that is why the space loss used is the loss at nadir). There are three antennas at each gateway and each of the antennas track a specific satellite in view. A 1 dB tracking loss is taken. The transmitted data rate is six times the actual data rate due to the beam hopping and TDD operation.

36

All the links are chip synchronous CDMA except for the user to satellite links. Orthogonal codes are used so there is literally no self-interference in those links. There are 128 orthogonal codes based on Walsh functions. There is very little self-interference on the gateway to satellite uplink. The gateways and satellites (in the OBP case) use convolutional, rate 1/2, constraint length 9 encoders with interleaving. The user unit uses a convolutional, rate 1/3, constraint length 9 encoder. There is a 1.3 dB interference margin required for the fading, blockage and power control. In the bent-pipe user to gateway link, a 1.0 dB modem/Doppler loss is taken due to the Doppler estimation required by the gateway. This is much more accurate in the OBP user to satellite link because the satellite can make faster and more accurate Doppler estimations.

Figure 6 Bent-Pipe Link Budget

	ES to Satellite	User to Satellite		Satellite to ES	Satellite to User
Frequency	8530.0	1825.0 MHz	Frequency	5205.0	1825.0 MHz
RF Power/User (peak)	0.2	6 Watts	RF Power/sub-band (peak)	1.2	18 Watts
	-7.0	7.8 dBW		0.8	12.0 dBW
Power Loss	-1.0	-1.0 dB	Power Loss	-1.5	-2.2 dB
Antenna Gain	40.2	2.0 dBi	S/C Ant. Gain (isoflux)	1.0	8.5 dBi
EIRP	32.2	8.8 dBW	EIRP	0.3	18.3 dBW
No. Users, u	25.0	25.0	Pilot Power	0.0	-0.2 dB
Voice Duty Cycle	37.5	37.5 %	No. Users & Interferers, u	27.0	25.0
-10*log (u*d)	-9.7	-9.7 dB	Voice Duty Cycle, d	37.5	37.5 %
			-10*log (u*d)	-10.1	-9.7 dB
			EIRP/channel	-9.8	8.4 dBW
Satellite Altitude-750 NM	1389.0	1389.0 km	Satellite Altitude-750 NM	1389.0	1389.0 km
Elevation Angle	90.0	90.0 degrees	Elevation Angle	90.0	90.0 degrees
Range	1389.0	1389.0 km	Range	1389.0	1389.0 km
Free Space Loss	-171.6	-159.5 dB	Free Space Loss	-169.7	-159.5 dB
RX Signal Strength	-139.4	-150.8	RX Signal Strength	-179.4	-153.1
Polarization Loss	-1.0	-0.5 dB	Polarization Loss	-1.0	-0.5 dB
Tracking Loss	-1.0	0.0 dB	Tracking Loss	-1.0	0.0 dB
S/C Ant. Gain (isoflux)	1.0	6.5 dBi	Antenna Gain	38.2	2.0 dBi
RX Line Losses	-1.2	-1.0 dB	RX Line Losses	-1.0	-0.5 dB
Data Rate, Rb	28800.0	28800.0 bps	Data Rate, Rb	28800.0	28800.0 bps
-10*log(Rb)	-44.6	-44.6 dB/Hz	-10*log(Rb)	-44.6	-44.6 dB/Hz
Eb	-186.2	-190.4 dBW/Hz	Eb	-188.8	-196.7 dBW/Hz
LNA Temp.	170.0	75.0 degree K	LNA Temp.	85.0	75.0 degree K
Ant. Noise	290.0	290.0 degree K	Ant. Noise	150.0	200.0 degree K
Total Thermal Noise	460.0	365.0 degree K	Total Thermal Noise	174.8	264.0 degree K
Thermal Noise Density, No	-202.0	-203.0 dBW/Hz	Thermal Noise Density, No	-206.1	-204.4 dBW/Hz
Single Interferer Power	-141.8	-145.8 dBW	Single Interferer Power	-144.2	-152.1 dBW
Equivalent # of Interferers	0.0	2.0	Intra/Inter Beam Interference	-134.3	-149.4 dBW
Self Interference	-155.0	-147.0 dBW	Inter Satellite Interference	0.0	25.0 %
Other System Interference	-155.0	-155.0 dBW	Self Interference	-134.3	-145.9 dBW
Total Interference	-152.0	-148.4 dBW	Other System Interference	-155.0	-155.0 dBW
Spreading Bandwidth	1.25	1.25 MHz	Total Interference	-134.3	-145.4 dBW
-10*log(spreading BW)	-61.0	-61.0 dB/Hz	Spreading Bandwidth	1.25	1.25 MHz
Pseudo-Noise Density, Io	-213.0	-207.3 dBW/Hz	-10*log(spreading BW)	-61.0	-61.0 dB/Hz
			Pseudo-Noise Density, Io	-195.3	-206.3 dBW/Hz
Thermal, Eb/No	15.8	12.6 dB	Thermal, Eb/No	17.3	7.7 dB
Pseudo Noise, Eb/Io	26.8	17.0 dB	Pseudo Noise, Eb/Io	9.0	9.6 dB
Uplink Eb/(No+Io)	15.4	11.3 dB	Downlink Eb/(No+Io)	6.1	5.5 dB
			C/I (Intermodulation)	23.0	15.0 dB
			Overall Thermal, Eb/No	11.3	7.0 dB
			Overall Interference, Eb/Io	6.4	9.6 dB
			Overall Eb/(No+Io+M)	5.2	4.7 dB
Number of L-Band Beams	6		Modem Imprm/Doppler Loss	-1.0	-0.5 dB
Number of SubBands	13		Eb/(No+Io) Required	3.0	3.5 dB
C-Band Amplifier Efficiency	18 %		Inter/Power Control Margin	1.3	1.3 dB
L-Band Amplifier Efficiency	35 %		Thermal Noise Margin	1.0	1.0 dB
Number of Beams Transmitting at Once	2		DC POWER		681.0 Watts
			TOTAL Mobile BW		16.50 MHz
			TOTAL USERS		1950.0

Figure 7 OBP Link Budget

	ES to Satellite	User to Satellite		Satellite to ES	Satellite to User
Frequency	8530.0	1825.0 MHz	Frequency	5205.0	1825.0 MHz
RF Power/User (peak)	0.1	3.5 Watts	RF Power/channel (peak)	0.1	15 Watts
	-10.0	5.4 dBW		-10.0	11.8 dBW
Power Loss	-1.0	-1.0 dB	Power Loss	-1.5	-2.2 dB
Antenna Gain	40.2	2.0 dBi	S/C Ant. Gain (isoflux)	1.0	8.5 dBi
EIRP	29.2	6.4 dBW	EIRP	-10.5	16.1 dBW
No. Users, u	25.0	25.0	Pilot Power	0.2	-0.2 dB
Voice Duty Cycle	37.5	37.5 %	No. Users & Interferers, u	25.0	25.0
-10*log (u*d)	-9.7	-9.7 dB	Voice Duty Cycle, d	37.5	37.5 %
			-10*log (u*d)	-9.7	-9.7 dB
			EIRP/channel	-9.7	6.1 dBW
Satellite Altitude-750 NM	1389.0	1389.0 km	Satellite Altitude-750 NM	1389.0	1389.0 km
Elevation Angle	90.0	90.0 degrees	Elevation Angle	90.0	90.0 degrees
Range	1389.0	1389.0 km	Range	1389.0	1389.0 km
Free Space Loss	-171.6	-159.5 dB	Free Space Loss	-169.7	-159.5 dB
RX Signal Strength	-142.4	-153.1	RX Signal Strength	-180.1	-153.4
Polarization Loss	-1.0	-0.5 dB	Polarization Loss	-1.0	-0.5 dB
Tracking Loss	-1.0	0.0 dB	Tracking Loss	-1.0	0.0 dB
S/C Ant. Gain (isoflux)	1.0	6.5 dBi	Antenna Gain	38.2	2.0 dBi
RX Line Losses	-1.2	-1.0 dB	RX Line Losses	-1.0	-0.5 dB
Data Rate, Rb	28800.0	28800.0 bps	Data Rate, Rb	28800.0	28800.0 bps
-10*log(Rb)	-44.6	-44.6 dB/Hz	-10*log(Rb)	-44.6	-44.6 dB/Hz
Eb	-189.2	-192.7 dBW/Hz	Eb	-189.5	-197.0 dBW/Hz
LNA Temp.	170.0	75.0 degree K	LNA Temp.	85.0	75.0 degree K
Ant. Noise	290.0	290.0 degree K	Ant. Noise	150.0	200.0 degree K
Total Thermal Noise	460.0	365.0 degree K	Total Thermal Noise	178.8	264.0 degree K
Thermal Noise Density, No	-202.0	-203.0 dBW/Hz	Thermal Noise Density, No	-206.1	-204.4 dBW/Hz
Single Interferer Power	-144.8	-148.1 dBW	Single Interferer Power	-154.9	-152.4 dBW
Intra/Inter Beam Interference	0.0	26.0 dBW	Intra/Inter Beam Interference	-152.1	-148.7 dBW
Self Interference	-155.0	-138.7 dBW	Inter Satellite Interference	0.0	25.0 %
Other System Interference	-155.0	-155.0 dBW	Self Interference	-152.1	-148.1 dBW
Total Interference	-152.0	-138.1 dBW	Other System Interference	-155.0	-155.0 dBW
Spreading Bandwidth	1.25	1.25 MHz	Total Interference	-150.3	-145.6 dBW
-10*log(spreading BW)	-61.0	-61.0 dB/Hz	Spreading Bandwidth	1.25	1.25 MHz
Pseudo-Noise Density, Io	-213.0	-199.1 dBW/Hz	-10*log(spreading BW)	-61.0	-61.0 dB/Hz
			Pseudo-Noise Density, Io	-211.3	-206.6 dBW/Hz
Thermal, Eb/No	12.8	10.3 dB	Thermal, Eb/No	6.6	7.4 dB
Pseudo Noise, Eb/Io	23.7	6.4 dB	Pseudo Noise, Eb/Io	11.8	9.6 dB
C/I (Intermodulation)	50.0	50.0 dB	C/I (Intermodulation)	23.0	15.0 dB
Uplink Eb/(No+Io-M)	12.4	4.9 dB	Downlink Eb/(No+Io-M)	5.4	4.9 dB
			Modem Imprm/Doppler Loss	-0.5	-0.5 dB
			Eb/(No+Io) Required	3.7	3.7 dB
			Inter/Power Control Margin	0.5	1.3 dB
			U/L Thermal Noise Margin	8.5	1.0 dB
Number of L-Band Beams	6		DC POWER		564.4 Watts
Number of SubBands	13		TOTAL Mobile BW		16.50 MHz
C-Band Amplifier Efficiency	18 %		TOTAL USERS		1950.0
L-Band Amplifier Efficiency	35 %				
Number of Beams Transmitting at Once	2				

For the bent-pipe links, the user to satellite has an additional 10% of the users due to interbeam and intersatellite interference. Also note that this additional power is used on the satellite to gateway downlink because it is a bent-pipe (all that is received is transmitted). The satellite to user downlink has 20% of intra-satellite interference and 25% inter-satellite interference. The satellite to gateway link is the limiting link due to the fact that the users within a channel are not chip-synchronous. This self-interference is the limiting factor for the bent-pipe transponder. The bent-pipe satellite can support 1950 simultaneous duplex calls and required 681 watts of DC power (while transmitting).

For the OBP links, the user to satellite path has an additional 10% interference along with the non-synchronous users in each channel. As in the bent-pipe case, this link is the limiting path. The satellite to user downlink and the satellite to gateway downlink have 20% if intra-satellite interference while the satellite to user path has an additional 25% due to inter-satellite interference. The OBP satellite can support 1950 simultaneous duplex calls while requiring only 565 watts of DC power for the transmitters. This does not take into account the additional power required for the OBP digital equipment and control.

However, the OBP satellite is not limited to 1950 calls. The OBP satellite can serve up to 2300 duplex calls until the user to satellite path self-interference limits the number of users (the required communications payload power is 720 watts).

Comparison

The link budgets show that the OBP transponder is more RF power efficient than the bent-pipe transponder (0.29 watts/call versus 0.35 watts/call for the bent-pipe transponder, an 18% savings). However, the power required for the digital equipment of the OBP transponder is not taken into account. For 1950 duplex calls, 2000 to 3900 modems are required depending on the design of the satellite-to-gateway links (only 2000 modems if TDMA is used for the satellite-to-gateways links). For the OBP transponder to be more power efficient than the bent-pipe transponder, the D.C. power of modem needs to be less than 0.06 watts each. It is not believed that this can be achieved until after the mid 1990's. Currently, this type of CDMA modem requires a power consumption of approximately 0.5 watts. This raises the call power from 0.29 watts/call to over 1.2 watts/call. This higher call power is 368% more than the bent-pipe transponder. Therefore, the bent-pipe transponder is more power efficient than the OBP transponder when the power for the OBP equipment is added.

There is more volume required for the OBP payload than the bent-pipe payload and the OBP payload dissipates more power. This extra dissipation could present a problem for small spacecraft. The OBP satellite will be more complex than the bent-pipe satellite, requires more design effort, and the risk is higher.

The OBP transponder does have advantages. The OBP transponder makes better use of the satellite's EIRP by only transmitting the signals of the calls going through **that** satellite (whereas a bent-pipe satellite transmits everything it receives). This gives an increase in call capacity for a given region due to decreased interference and the OBP satellite is able to make better use of satellite double coverage. For CDMA operation, the OBP transponder will have better power control. Also with an OBP transponder, the call set-up procedure can be moved to the satellite.

The cost and complexity of the OBP satellite gateways versus the bent-pipe satellite gateways depends on the design of the OBP gateway-to-satellite links. By using TDMA links, the OBP gateway costs might be higher than the bent-pipe solution. By using CDMA links, the OBP gateway costs might be lower than the bent-pipe gateway costs, but doubles the number of modems in the satellite from 1950 to 3900. The price of a space qualified CDMA modem is not available today since no qualified unit exists. However, it is obvious a cost increase will be incurred with an OBP transponder. For example, 3900 CDMA modems reduced to a single VLSI chip when space qualified would be on the order of $500 each. At this cost, the price of a satellite chip set is nearly two million dollars not including other OBP equipment, integration, and test. A 48 satellite set would incur additional costs of $100 million not including the additional satellite cost of the increased power requirements.

Another decisive factor is the non-standardization of the cellular systems in the world today. The GLOBALSTAR system must be compatible with many types of mobile operations. Europe, the United States and Japan all have different standards. A bent-pipe transponder allows for these different standards to be used with the satellite. There is not enough flexibility in an OBP transponder to handle the different standards and protocols.

Conclusions

The first generation GLOBALSTAR satellite will be bent-pipe due to the flexibility of the transponder to handle different signal formats. A bent-pipe transponder offers inexpensive capacity to users without the prohibitive research and development, power and cost requirements that would be incurred with an OBP solution. The bent-pipe communications subsystem is a classic repeater which uses existing satellite communications techniques and many off-the-shelf parts. This keeps the nonrecurring research and development costs low as well as keeping the satellite equipment and testing costs low. Table 1 give an overview of the comparison between OBP and bent-pipe.

Table 1
Transponder Comparison

	OBP	Bent-Pipe
Cost/Channel		X
Signal flexibility		X
Payload Complexity		x
Capacity	x	
Use of Multi-coverage	X	
Payload Power		X
Payload Weight		x
Payload Volume		x
Thermal		x
Risk		x
R&D Required		x
Gateway Cost	-	-
Signal Quality†	x	

† at maximum capacity

The return link self-interference due to the non-synchronous CDMA operation limits the capacity of both the bent-pipe and OBP transponders. The OBP transponder is limited to 2300 simultaneous duplex calls while the bent-pipe transponder is limited to 1950 simultaneous duplex calls. However, the OBP transponder will use more power than the bent-pipe transponder, which is critical for this LEO satellite system.

Acknowledgements

References

1) J.G. Walker, **Continuous Whole-Earth Coverage by Circular-Orbit Satellite Patterns**, Royal Aircraft Establishment Technical Report 77044, 24 March, 1977.

Page intentionally left blank

Satellite Communications for the Next Generation Telecommunication Services and Networks

D. M. Chitre
COMSAT Laboratories
Clarksburg, MD 2087

Abstract

Satellite communications can play an important role in provisioning the next-generation telecommunication services and networks, provided the protocols specifying these services and networks are satellite-compatible and the satellite subnetworks, consisting of earth stations interconnected by the processor and the switch on board the satellite, interwork effectively with the terrestrial networks. This paper discusses the specific parameters and procedures of frame relay and broadband integrated services digital network (B-ISDN) protocols which are impacted by a satellite delay. Congestion and resource management functions for frame relay and B-ISDN are discussed in detail, describing the division of these functions between earth stations and on board the satellite. Specific on-board and ground functions are identified as potential candidates for their implementation via neural network technology.

1. Introduction

Next-generation telecommunication networks are currently being shaped as a result of the rapid development of some key concepts and the associated detailed standards for integrated services digital networks (ISDN), both nationally (American National Standards Institute [ANSI] Committee T1) and internationally (International Telegraphy and Telephony Consultative Committee [CCITT] Study Groups XVIII, XI, and I). The ISDN defines a worldwide communications environment encompassing all digital information, not only voice and data, but also facsimile, videophone, videoconferencing, interactive cmputer-aided design/computer-aided manufacturing (CAD/CAM) and any other type of information that can be digitized. There is a symbiotic relationship between the emerging ISDN environment and satellite communications. Digital communications channels provided by various domestic and international satellite networks can form a digital transmission backbone for the establishment of worldwide ISDN. The well-specified ISDN standards on interfaces, signaling procedures, service definitions, and other operational and architecture features will provide a common framework for developing satellite network configurations offering a range of services. The key features of a satellite communications system, namely, agility of communication bandwidth, multipoint/broadcast nature of satellite channels, global coverage and mobility of very small aperture terminal/mobile satellite (VSAT/MSAT) terminals play a very crucial role in the future development of ISDN, intelligent networks, Future Public Land-Mobile Telecommunications System (FPLMTS), and finally, the Universal Personal Telecommunications (UPT) service. UPT is a telecommunications service which will enable the user to establish and receive calls and services on the basis of a unique personal telecommunication number (PTN) across multiple networks at any user-network access whether fixed, movable, or mobile irrespective of the geographical area. The worldwide satellite networks along with VSAT and the mobile satellite networks will go long ways to make UPT service feasible.

Satellite communications can play a significant role in the future national and international telecommunication networks provided the satellite subnetworks are closely integrated with the terrestrial networks. The specific integration issues are discussed in this paper within the context of the networking (and underlying protocol) standards which are going to be the fundamental underpinnings of future terrestrial networks.

This paper will focus on emerging standards in two areas, frame relay and broadband ISDN (B-ISDN); and discusses, in detail, their impact on the satellite networks with on-board processing and switching.

2. Frame Relay

2.1 Service and Protocol Description

Frame relay is a new ISDN packet mode bearer service for data communications at access speeds of up to 2.048 Mbit/s [I.122, I.233].* This bearer service provides the order-preserving bidirectional transfer of layer 2 frames from the source user-to-network interface (S or T ISDN reference point) to the destination network-to-user interface (another S or T ISDN reference point). The data units (called frames) are routed through the network on the basis of an attached label termed, data link connection identifier (DLCI). The DLCI identifies a virtual connection on a bearer channel (i.e., D, B, or H) at a user-to-network or network-to-network interface (UNI or NNI). The major characteristics of this service are the logical out-of-band call control using protocol procedures that are integrated across all telecommunications services and the statistical multiplexing of different user data streams (via DLCI) at the link layer in the user plane.

Flow-control and error-recovery functions are performed on an end-to-end basis by a user-selectable, end-to-end protocol. LAPF (Q.922) is currently being developed for use as one of these end-to-end protocols. The link layer parameters chosen for LAPF are important in determining the effectiveness of the satellite networks providing frame relay service. The parameter for the retransmission timer, T200, with a default value of 1.5 s, can accommodate a one-hop satellite delay. The other parameter, k, (the maximum number of outstanding frames) can have values ranging from 1 to 127. The default value, as currently specified for a 64-kbit/s channel, is seven which is too low. It will result in the inefficient operation of the frame relay service over satellite channels. The value of k should be negotiated to a higher value (such as 40 for a 128-octect frame size) at the call set-up time via the XID procedure described in Appendix III of Q.922.

A subset of LAPF, corresponding to the data link core sublayer is used to support the frame relaying bearer service. The network does not support any procedures above these core functions of Q.922; such as acknowledging frames (within the network), or keeping the sequence numbers. The core functions are:

- frame delimiting, alignment, and transparency
- frame multiplexing, and demultiplexing using the address field
- inspection of the frame to ensure that it consists of the integer number of octets prior to zero bit insertion or following zero bit extraction
- inspection of the frame to ensure that it is neither too long nor too short
- detection of transmission errors
- congestion control functions.

---

* ISDN Recommendations referred to in this paper can be found either in the CCITT Blue Books, ITU, Geneva 1988 or as draft recommendations output from the CCITT Study Group XVIII Meeting in Matsuyama, Japan, 11/26/90–12/7/90 and XI Meeting in Geneva, Switzerland, April 1991.

The address field of the frame consists of at least two octects, containing a DLCI identifying a virtual connection on a bearer channel. The field variables in the address field for the congestion management are as follows.

a. Forward explicit congestion notification (FECN):

This bit can be used by a frame relaying network node to notify the user that the congestion-avoidance procedures should be initiated where applicable for traffic in the direction of the frame carrying the FECN indication. This bit is set to 1 to indicate to the receiving-end system that the frames it receives have encountered congested resources. This indication can be used by the destination-controlled transmitter rate adjustment.

b. Backward explicit congestion notification (BECN):

This bit can be set by a congested network to notify the user that the congestion-avoidance procedure should be initiated, where applicable, for traffic in the opposite direction of the frame carrying the BECN indicator. This bit is set to 1 to indicates to the receiving-end system that the frames it transmits may encounter congested resources. This indication can be used by the source-controlled transmitter rate adjustment.

While setting of the above bits by the network or user is optional, the network is not allowed to clear (set to 0) these bits. Networks that do not provide FECN or BECN will pass this bit unchanged.

c. Discard eligibility (DE) indicator:

This bit, if used, is set to 1 to indicate a request that a frame should be discarded in preference to other frames in a congestion situation. The setting of this bit by the network or user is optional. A network will never clear (set to 0) this bit. Networks that do not provide DE will pass this bit unchanged. Networks are not constrained to discard frames with DE =1 in the presence of congestion.

2.2 Congestion Management

Congestion in the user plane occurs when the traffic arriving at a resource exceeds the capacity of the network. In the frame relay networks, congestion can arise due to users offering in excess of the committed traffic, or due to coincidental peak traffic demands, or, possibly, due to equipment failure and the consequent degraded network capabilities. In these networks, congestion control is achieved via both the congestion-avoidance mechanisms and congestion-recovery mechanisms. The former is used at the onset of congestion via the explicit congestion notifications. For destination-controlled transmitters, the FECN bit is set in the appropriate frames. For source-controlled transmitters, the BECN bit is set in frames transported in the reverse direction (i.e., towards the transmitter). Alternatively, a consolidated link layer management (CLLM) message can be generated providing reverse notification for one or more DLCIs within a single frame. The CLLM is sent on the layer management DLCI in the user plane.

Congestion-recovery mechanisms are used to prevent network collapse in the event of severe congestion. Implicit congestion detection and end-user responses are defined in Q.922 to recover from congestion.

One of the important issues for satellite compatibility in frame relay networks is the implementation of a congestion-recovery mechanism that works efficiently with large propagation delays. Congestion recovery in frame relaying will be done by adjusting the sizes of the LAPF windows that control the number of outstanding, unacknowledged frames. In general, these windows must be relatively large when satellite links are employed in order to account for the propagation time. If the window sizes are too small, a transmitter will constantly be stopping transmission to wait for acknowledgment of previous frames. This results in inefficient utilization of the satellite channel, along with increased delay and decreased throughput for the user.

Appendix I of Draft Recommendation Q.922 discusses the use of dynamic window size to respond to network congestion and describes the congestion-control algorithm.

The algorithm modifies the transmitting-data-link layer entity's transmit window when the congestion is first detected and again as congestion decreases. The congestion-control algorithm is triggered by the loss of I frames. When the data-link layer detects this loss, either by the reception of an REJ frame, or the confirmed T200 expiration event, it invokes the dynamic window algorithm and reduces its transmit window size to a fraction of its original size. The transmit window size is gradually increased until it returns to its original value, k, its value in the absence of congestion.

For satisfactory operation of LAPF (Q.922) over a 64-kbit/s satellite link, the value of k during the absence of congestion should be as large as 40 for a 128-byte frame size (and even larger for higher-rate channels, H_0, H_{11}, and H_{12}). Thus, a reduction in the window size, for example, from 40 to 10 due to the loss of one frame and the gradual increase over the next several round-trip delays will cause considerable reduction in throughput. Notice also that the frame loss is not necessarily due to congestion, but could be due to a bit error.

A somewhat different congestion control strategy, where the transmit window is reduced gradually, for example, by a factor of two at each step, and a more rapid increase will considerably alleviate the problems for the satellite operation without impacting the terrestrial operation in most cases.

Appendix I of Q.922 also discusses the user response on the receipt of the FECN, the BECN and the CLLM.

2.3 Frame Relay Functions For Satellite Subnetworks

Specific functions within the satellite subnetwork are identified here for the support of the frame relay bearer services. These functions are largely independent of the specific details of the satellite network architecture. A generic model of the satellite subnetwork within a frame relaying network is considered. The nodes of the satellite subnetwork consist of distributed earth stations and an on-board packet processor and a switch interconnecting up-link and down-link beams. The interfaces to the terrestrial subnetworks are at the earth stations which perform the frame relay core functions. The on-board processor also implements all the frame relay core functions described in Subsection 2.1. However, the specific congestion-control functions (and algorithms) undertaken at the earth stations and on board differ based on the amount of information available and the complexity of the algorithm.

2.3.1 Functions at the Earth Stations

An earth station of the satellite subnetwork can be an ingress node, an egress node, or just an intermediary network node within a frame relaying network. The following functions are necessary for imminent congestion determination and subsequent proper response.

- traffic monitoring
- congestion determination via thresholding
- FECN, BECN/CLLM implementation
- discard eligibility flag insertion.

Each earth station acting as a frame node monitors the frame traffic on each DLCI connection entering into and going out of the satellite subnetwork. The offered traffic is compared with the committed level at the call setup. The specific traffic parameters are the committed burst size (Bc), excess burst size (Be), and committed information rate (CIR). These parameters are measured over a certain computed time interval (Tc).

The congestion determination can be done by computing the average queue length for each outgoing link, including the up-link to the satellite. Appropriate thresholds T1 and T2 are set such that when the average queue size is greater than T2, the link is considered to be in the state of congestion. Beginning at that time, and continuing until average queue size falls below the threshold, T1, the link is declared to be congested.

Whenever, an outgoing link from the earth station is considered congested, the FECN bit is set to 1 on all outgoing frames on that link. Appropriate CLLM messages are generated for transmission on all of the incoming links contributing the traffic to the outgoing congested link. CLLM can be replaced by frames with BECN bit set to 1, if there is current reverse traffic on those links.

Frames offered in excess of the committed traffic negotiated at the call setup are marked as discard eligible at the earth station before transmitting them forward. However, if the outgoing link is in the congested state, the frames marked discard eligible and buffered for transmission on that link are dropped.

2.3.2 Functions On Board the Satellite

The on-board processor performs the core functions described in Subsection 2.1 and routes the frames based on the mapping tables relating the DLCI number and the down-link channel.

However, the congestion-management functions on board the satellite should be simpler than at the earth stations. Specifically, the detailed monitoring of all the separate DLCI offered traffic and its comparison with the committed level may not be performed on board. The frame traffic buffered for each outgoing down-link is monitored. However, the threshold method of determining a mild congested state of a link may not be cost-effective on board the satellite since it will take some time before the users respond, over the satellite link, to the FECN, BECN, or CLLM messages to relieve the congestion. In the interim, there could be considerable amount of traffic in the pipeline worsening the congestion. On the other hand, setting the threshold value too low will underutilize satellite channels.

A predictive method which does not depend upon implementing a very complex algorithm on board the satellite is, thus, highly desirable. A suitable neural network (NN) implementation scheme could, indeed, be such a method. A feed-forward back-propagation network, with the input to the NN consisting of the traffic pattern on a particular down-link channel, can be trained to give the correct output. The output will be a binary decision whether that link is

entering a congested state or not. To train the NN, a number of traffic patterns can be simulated and their impact on the future state of the link observed. With an appropriate definition for the link congestion, different patterns can then be correlated with the predictive behavior of the future state of the link.

The congestion detection based on the NN output can trigger the generation of appropriate FECN, BECN, or CLLM messages and the dropping of frames with discard eligibility set to 1.

It should be noted that the NN technique can also be used at the earth stations for the detection of oncoming congestion.

3. B-ISDN

3.1 B-ISDN Protocols

Although the CCITT work on the formulation of broadband ISDN (B-ISDN) recommendations is still in its early stages, several issues concerning broadband services and network capabilities and requirements have already been agreed upon. Asynchronous transfer mode (ATM) is the proposed transport technique for the B-ISDN. ATM is a packet mode information method which uses fixed-size packets called cells. The cells are statistically multiplexed and are identified as belonging to a particular logical connection by the virtual channel identifier (VCI) that is carried as a label in the header of each cell. A virtual path, identified by a virtual path identifier (VPI), is a grouping of virtual channels. ATM offers the flexibility to support a wide variety of service types (voice, data, or video) and provides efficiency by statistically multiplexing possible bursty traffic. There is a lot of similarity between frame relay and ATM. The major differences are that frames are of variable size while cells are fixed size (48 bytes of user data and 5 bytes of header); and ATM cell header has more network functionality (I.121, I.150, I.361).

ATM adaptation layer (AAL) protocols on top of the ATM layer are being specified for different types of services. Recommendation I.363 defines an assured mode operation for data transfer for a Type 4 (or Type 3) AAL convergence protocol. For the high bit rates envisaged for B-ISDN, the choice of the protocol for the assured mode operation needs to be made after careful analysis.

Two recovery strategies are likely candidates for the Type 4 (or Type 3) ATM adaptation layer convergence protocol for assured operation. They are:

- *go-back-N (GBN)*—retransmit all messages from missing sequence number, and
- *selective retransmission (SR)*—retransmit only messages corresponding to missing sequence number.

In the GBN method, the AAL messages which are encoded for error detection are transmitted sequentially and the acknowledgments from the receiver arrive after a round-trip delay. During this delay, which is the time between the transmission of the AAL message and the receipt of its acknowledgment, N-1 other messages are also transmitted. When the receiver gets an erroneous message or an out-of-sequence message, a negative acknowledgment (i.e., an REJ message) is sent by the receiver. When a negative acknowledgment is received, the transmitter stops sending new messages, backs up to the negatively-acknowledged message and retransmits it and all subsequent messages, thus giving rise to spurious retransmissions of at least N-1 messages.

At high link speeds, such as 45 Mbit/s, this phenomenon causes a severe degradation in both throughput and delay even at low bit error rates (BER) and the performance deteriorates

very rapidly as the BER or message-loss rate increases. Furthermore, if the loss of a message is due to mild congestion, the spurious retransmissions will aggravate the congestion giving rise to still more messages being lost causing more spurious retransmissions and so on.

Figure 1 shows the degradation in throughput for the GBN protocol operating over a 45-Mbit/s satellite connection. (Note that the bit error ratio is really an effective cumulative bit error ratio arising out of line errors and packet losses.) Figure 2 is the corresponding curve over a 45-Mbit/s terrestrial connection. The selective retransmission curves are based on simulations of the protocol and the GBN curves are based on analytical results.
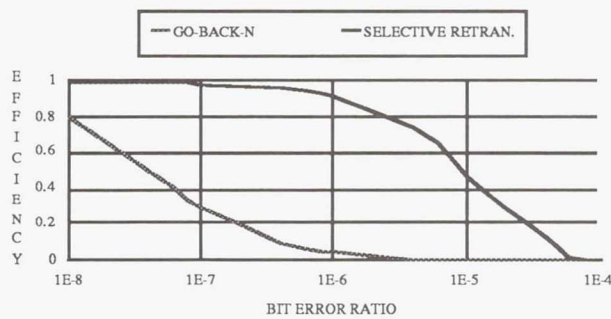
FIGURE 1. *Efficiency of a 45 Mbit/s Satellite Link for a Frame Size of 4,096 Octets*
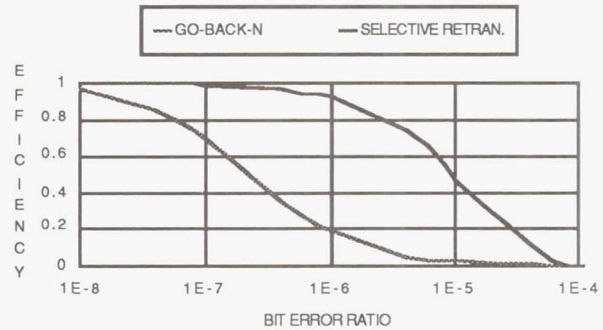
FIGURE 2 *Efficiency of a 45 Mbit/s Terrestrial Link for a Frame Size of 4096 Octets*

The GBN strategy, which has a certain simplicity for implementation, however, performs very poorly at high speeds which are likely to be encountered by the AAL convergence protocol. It will thus be desirable to remove the option of GBN error-recovery procedure and introduce new generation of simple-to-implement selective-retransmission protocols.

3.2 Traffic Control and Resource Management

The advantages of the ATM, however, can be seriously hindered if effective traffic and congestion control techniques are not implemented by the network. At the high-speed cell transport level, the congestion control techniques proposed in frame relaying networks are not adequate. Additional mechanisms like admission control (called acceptance/rejection) need to be used in conjunction with bandwidth enforcement and flow control.

Since the B-ISDN, using the ATM technique, is designed to transport a range of traffic classes with a widely varying traffic and quality of service (QOS) requirements, it is essential to have several levels of traffic control capabilities such as:

- connection admission control (CAC)
- usage parameter control (UPC)
- network parameter control (NPC)
- priority control
- congestion control.

Connection admission control is defined as the set of actions taken by the network at the call setup phase (or during call re-negotiation phase) in order to establish whether a virtual channel connection (VCC) or a virtual path connection (VPC) can be accepted or rejected. On the basis of the connection admission control outcome in an ATM based network, a connection request for a

47

given call is accepted only when sufficient resources are available to establish the call through the whole network at its required QOS and to maintain the agreed QOS of existing calls.

UPC and NPC perform similar functions at different interfaces such as UNI and NNI.

UPC/NPC is defined as the set of actions taken by the network to monitor and control (user) traffic in terms of traffic volume and cell routing validity. Its main purpose is to protect network resources from malicious as well as unintentional misbehavior, which can affect the QOS of other already established connections by detecting violations of negotiated parameters. The possible parameters could be cell-peak rate, average rate, burstiness, or peak duration.

Priority control is based on the fact that the user is allowed different priority traffic flows by using the cell loss priority (CLP) bit in the ATM header (I.150).

Congestion control can work in two modes, preventive and reactive, depending upon the state of the network. In the preventive mode, connection admission control takes into account the current load on the network and rejects the call request. In the reactive mode, techniques based on the discarding of cells with the high CLP value or cells carrying violation tags, coupled with congestion notification can be used.

3.2.1 Functions at the Earth Stations

Figure 3 illustrates the B-ISDN satellite subnetwork (enclosed in the ellipse). The B-ISDN controller at an earth station performs a number of functions including traffic monitoring and UPC (if the earth station is an ingress node) or NPC (if the earth station is an intermediary node), connection admission control, violation tagging for the cells, congestion prediction/detection, congestion notification, and discarding of cells.
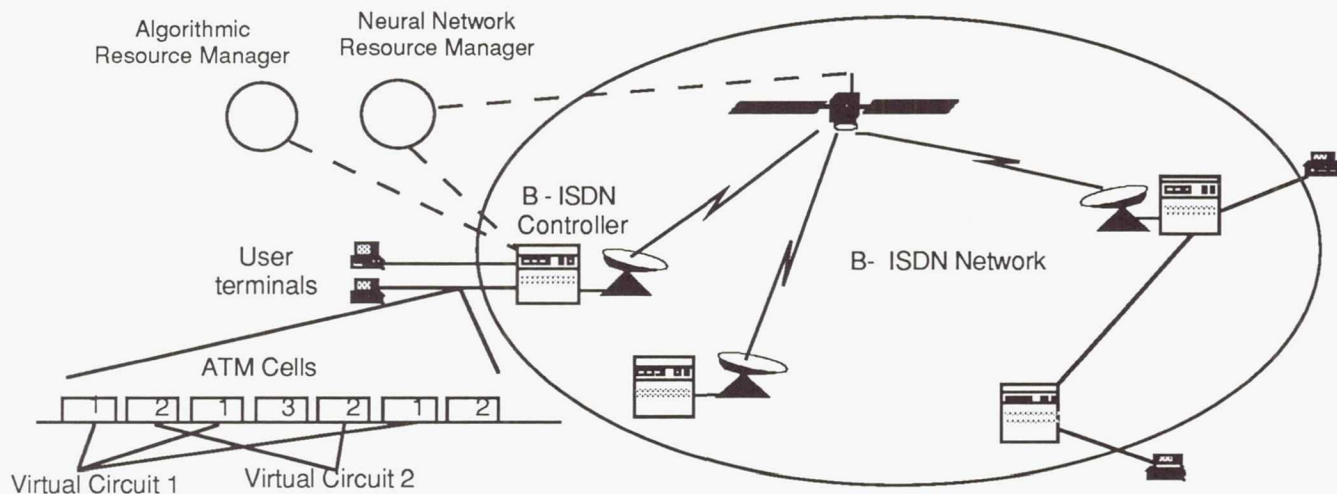
Figure 3: Satellite B-ISDN Network With Onboard Processing and Switching

The ATM cell traffic is monitored for every VPC/VCC to find if it exceeds the negotiated value at the call setup for that connection. The cells in violation can be tagged either using the one bit reserved field or perhaps setting the CLP bit to 1.

48

Congestion for any outgoing link (including the up-link to the satellite) can be detected either by a classical algorithm (through setting buffer thresholds) or more predictive neural network models can be used. Two different feed-forward back-propagation networks can be considered. The first one will implement the connection admission control. It will decide whether a particular call request, with appropriate traffic descriptors, should be accepted or rejected during a particular state of the network. Notice that the call may have to be rejected even if the current state of the network is not congested. (The acceptance of the call could lead to congestion characterized by exceeding certain levels of cell loss ratio.) The second neural network will predict the future state of the network for the existing calls. In that case, the congestion could arise due to coincidental peak traffics from different sources or the degradation of network resources. The congestion prediction decision of the second neural network can be used to discard cells with violation tags or cells with the high CLP bit and for sending explicit congestion messages as operations, administrations, and management (OAM) cells to the appropriate VPI/VCI cell sources.

3.2.2 Functions On Board the Satellite

The on-board processor routes the ATM cells either individually or as satellite virtual packets (SVP) which consist of a number of ATM cells (with the same CLP) destined for the same down-link channel. (Such concatenation could have been done at an earth station.) The detail monitoring of traffic on individual VPC/VCC may not be performed on board the satellite. However, the ATM cell or SVP traffic buffered for each down-link channel is monitored. The two NNs described in Subsection 3.2.1 can be implemented on board for CAC and detecting the oncoming congestion. Based on the output of these NNs, the NN resource manager, shown in Figure 3, can make a decision regarding the acceptance or rejection of the new call and discarding the cells with violation tags or cells with the CLP bit set to 1. Upon detection of oncoming congestion, explicit congestion notification messages can be transmitted to earth stations contributing to the congestion. The earth stations having monitored the individual VPC/VCC traffic can then generate congestion indication messages for appropriate sources of corresponding VPC/VCC. The sources can then implement suitable flow control.

Notice that for B-ISDN high-speed traffic, the congestion and flow control based on the predictive, but simple to implement, NN on board the satellite is very important to alleviate the delay experienced by the notification messages over the satellite links.

4. Conclusions

Frame relay and B-ISDN services can indeed be provided very efficiently via a satellite network with on-board processing and switching. However, the future development of protocols for these services need to be properly shaped. Finally, the application of NN technology on the ground and on board can be brought to bear upon the congestion and resource management techniques to make optimum use of the satellite resources.

Page intentionally left blank

Satellite B-ISDN Traffic Analysis[1]

Dong-Jye Shyy and Thomas Inukai

COMSAT Laboratories
Clarksburg, MD 20871

Summary

This paper reports on the impact of asynchronous transfer mode (ATM) traffic on the advanced satellite broadband integrated services digital network (B-ISDN) with on-board processing. Simulation models have been built to analyze the cell transfer performance through the statistical multiplexer at the earth station and the fast packet switch at the satellite. The effectiveness of ground ATM cell preprocessing has been established, as well as the performance of several schemes for improving the down-link beam utilization when the space segment employs a fast packet switch.

1. Introduction

With its flexibility in bandwidth allocation for various traffic types, channel structures, and synchronization rates, B-ISDN is a promising technique to accommodate diverse current and future services and traffic. Potential applications for satellite B-ISDN include video distribution services, live newscast/broadcast, science data distribution, supercomputer networking, private network, and trunking. Providing these services necessitates the development of new network architectures, advanced on-board processing, and special processing at the earth station.

Two types of transfer mode can accommodate B-ISDN services: circuit switching and fast packet switching. Fast packet switching is more effective than circuit switching in providing services for multirate, multimedia, and bursty data due to its statistical multiplexing and inherent dynamic bandwidth allocation properties. Also the current trend in the transmission of telecommunications traffic is based on packet communications such as frame relay/frame switching, ATM, and consultative committee for space data systems (CCSDS). An on-board fast packet switch gives the satellite the flexibility to handle services with different traffic characteristics and requirements, and to provide higher transmission throughputs than circuit switching. Also, an on-board fast packet switch with multiple spot beam operation gives the satellite network an advantage over the terrestrial network in providing multicast services. In this paper, the earth station is assumed to be equipped with ATM interfaces and the space segment with a fast packet switch.

For ATM, the packet size is fixed and each packet, called a cell, consists of a 5-byte header and 48-byte information payload. In ATM cells, routing information consists of a 24-bit virtual path identifier and virtual channel identifier (VPI/VCI) at the user network interface (UNI) and a 28-bit VPI/VCI at the network node interface (NNI). To reuse VPI address space and avoid VPI retranslation on-board, the satellite virtual packet (SVP) concept is introduced. SVPs are created by appending a header, used only within the satellite network and containing a routing tag for the on-board switch, to one cell or a group of cells destined to the same down-link beam.

Since the satellite resources are bandwidth limited, the design focus is increase down-link beam utilization. To achieve this objective, one approach is to increase the throughput of the on-board fast packet switch. Different fast packet switch architectures have been proposed to improve the switch throughput [1]. Two fast packet switching architectures are considered in this paper: the input queueing fast packet switch with a nonblocking switching fabric and the output queueing fast packet switch with a nonblocking switching fabric.

For the input-queueing fast packet switch, the throughput is constrained by the head of line blocking problem. Three schemes to improve the throughput have been investigated: increasing the input buffer size, increasing the searching depth of the input queue to resolve the output contention, and increasing the switch speed. For the output-queueing fast packet switch, the incoming packets are not stored in the input buffers; the (banyan-type) switch must operate N times faster than the line speed to avoid the output contention problem, where N is the switch size.

The focus of this paper is performance analysis of the ATM cells (cell delay jitter and cell delay distribution) through the multiplexer at the earth station and the fast packet switch at the satellite using the SVP concept.

2. Satellite Virtual Packets

SVPs are created by appending a header to one cell or a group of cells destined to the same down link beam at the earth station (see Figure 1) for unified routing, control, and management. The header of the SVP is termed the satellite virtual label (SVL).

There are three possible formats for grouping the cells at the earth station: fixed size packets, variable size packets, or single cells. In this paper, the fixed size approach for SVPs has been adopted.

The necessary fields of the SVL are proposed to consist of the switch routing tag, sending earth station address, receiving earth station address, quality of service (QOS) field, control field, and cyclical redundancy check.

A larger SVP has a higher transmission capacity utilization. It also increases the packet interarrival time and the packet slot time, which decreases the speed requirement for on-board SVP processing. However, a larger SVP has a longer packetization delay, longer end-to-end delay, and worse delay jitter. It also results in a larger buffer requirement and increases the number of bits of the packet payload in error. In this paper, the performance of SVPs of different sizes are compared and analyzed in terms of delay through the satellite B-ISDN.

The tradeoff study between SVP processing and cell processing in the satellite B-ISDN has been omitted due to the page limitations.

3. Performance Analysis of SVP Ground Preprocessing and On-Board Processing

In this subsection, the cell delays through a multiplexer and a fast packet switch are collected from the simulation models to study the performance of SVP transmission through the satellite B-ISDN.

A brief description of the satellite B-ISDN simulation model (see Figure 2) is given below. The number of earth stations, up-link beams, and down-link beams is assumed to be the same as the size of the on-board switch. Each earth station is interfaced with five UNIs. These five input lines at the earth station have the same transmission rate of 155.52 Mbit/s. The up-link and down-link access schemes use time-division multiplexing (TDM). The link transmission rate is also 155.52 Mbit/s. The on-board switch speed is (speedup factor * 155.52 Mbit/s). The cell slot time is equal to 2.726E-6 sc when the transmission rate is 155.52 Mbit/s. In the figures, CDJ designates the cell delay jitter, u the link utilization, d the checking depth, and s the speedup factor of the switch.

3.1 SVP Transmission vs Cell Transmission at the Earth Station

Earth stations in the satellite B-ISDN are interfaced with different UNIs and network node interfaces (NNIs). To increase the transmission link efficiency and share the satellite transmission link among different users, the advanced earth station functions as a statistical multiplexer. At the earth station, it is assumed that the buffer size of the multiplexer is infinite; therefore, its cell loss ratio is zero.

The first set of results illustrates the effect of different SVP sizes on the cell delay performance and the buffer size requirements at the earth station for different transmission link utilizations and numbers of down-link beams.

Figure 3 shows the cell delay jitter of the statistical multiplexer at the earth station for different SVP sizes, output link utilizations, and numbers of down-link beams. As shown in this figure, for different sizes of SVPs, the increase in of the cell delay is proportional to the size of the SVPs. Based on these results, SVPs should be small if cell delay is an important QOS parameter. Also as shown in this figure, when the output link utilization is higher, the delay performance of SVP transmission improves while that of single-cell transmission degrades. That is, when the output link utilization is higher, the probability that an SVP will be filled with cells destined to the same down-link beam within a given time is also higher. Hence, the delay performance of the SVPs improves when the output link utilization is higher. Basically, the cell delay for the SVPs consists of three elements: the packetization delay (the time required to fill the SVP with cells), the waiting time in the queue, and the transmission time. From Figure 3, most of the delay for the SVPs occurs during the packetization process.

It can be seen from Figure 3 that the cell delay for SVP transmission is proportional to the number of down-link beams. As previously mentioned, the cell delay for SVP transmission is dominated by the packetization delay. If there are more down-link beams, the filling rate for each SVP destined to a different down-link beam is reduced when the output link utilization is kept constant.

In conclusion, the delay performance of cells for single-cell transmission is determined by the waiting time in the queue for transmission, i.e., the queueing delay. The delay performance of the cells for SVP transmission is largely determined by the packetization delay. This implies that, if the SVP transmission concept is used for the satellite network, the satellite network link must be operated at a very high utilization (above 80 percent) therefore, the on-board fast packet switch throughput must also be higher than 8 percent.

Figure 4 shows the cell delay probability mass function for different SVP sizes and link utilization when the number of down-link beams is 80. From the cell delay distribution, the buffer size requirement for the earth station to achieve a certain cell loss ratio can be derived. For example, for a 4-cell SVP and utilization equal to 0.6, the probability that the cell delay is 400 ms is about 10^{-4}. Hence, to have a cell loss ratio of 10^{-4}, the buffer size required for the statistical multiplexer is about 37 SVPs. With higher utilization, the cell delay distribution curves of the SVPs approach the cell delay distribution curves for single-cell transmission.

3.2 On-Board Fast Packet Switch Performance

For the on-board fast packet switch, since the mass and power are the constrained design factors for satellites, the buffer size of the switch must be finite. Also, since the satellite communications system is both power and bandwidth limited, the down-link beam resource must be used very efficiently. One way of increasing the down-link beam utilization is to increase the throughput of the switch. The on-board fast packet switch architectures are assumed to be banyan-type switching networks. The switching fabric is assumed to be unbuffered and point-to-point nonblocking. Since output blocking is unavoidable for a packet switch, it is assumed that packets are buffered either at the input ports or at the output ports.

3.2.1 Input Buffering

In this scheme, the packets are buffered at the input ports. The throughput of a switch with first-in first-out (FIFO) input queue is limited by the head of line blocking problem. It has been shown that the theoretical throughput for the input buffering nonblocking point-to-point switch with infinite buffer size is about 0.58 [2]. Head of line blocking is a side effect of output blocking. This is, if one packet at the head of queue cannot be transmitted due to output blocking, this packet hinders the delivery of the next packet in the queue due to the FCFS nature of the queue, even though the next packet can be transmitted to the destination without any blocking. Three schemes have been studied to improve the throughput of the switch. The first is to increase the buffer size. Intuitively, the larger the buffer size, the better the packet loss ratio, but the worse the delay performance. The second method is to use a non-FIFO queue. If the first packet is blocked due to output blocking, the scheduling algorithm will also examine the packets at the back of the first packet. The number of packets examined each time depends on the preset window size or the checking depth. For a normal FIFO operation, the checking depth is 1. The third method is to operate the switch at a speed higher than the link speed.

3.2.1.1 Increasing Buffer Size

The buffer requirement for the on-board switch is determined by measuring the cell delay distribution using infinite buffer size. From the delay distribution, the buffer size requirement for a specific cell loss ratio can be calculated as given in the example in Subsection 3.1. Figure 5 shows the cell delay jitter for different SVP sizes through the on-board switch. The delay degradation of a larger SVP through the fast packet switch is much less than that through the multiplexer at the earth station because, unlike the multiplexer, no packetization process is required at the switch. When the utilization is low, the single-cell delay performance is better than the SVP delay performance because the SVP transmission time through the switch is n times longer than the cell transmission time, where n is the size of the SVP. In general, under the same conditions, the packet delay through a switch for a packet of size n cells is n times larger than that for a single cell.

From Figure 5, when the link utilization is higher (approaching the 0.6 switch throughput of an 8 x 8 switch), the cell delay performance degrades much more quickly than the SVP delay performance. Eventually, the SVP delay performance is better than the single-cell delay performance because the process of formatting cells into SVPs at the

earth station disturbs the traffic pattern. Note that the traffic pattern coming to the simulated satellite network is assumed to be random. Under this scenario, the probability that n SVPs destined to the same destination beam arrive at the input port of the switch continuously is less than that of n cells, where n > 1. Hence, output contention of the on-board switch is reduced. This discovery is useful if the traffic coming to the earth station can be assumed to be random, as in the case of packet-switched traffic. In this case, the throughput of the on-board switch is increased by formatting the SVP at the earth station since the output contention problem of the switch is reduced.

The cell delay distribution for different SVP sizes and different link utilizations us is shown in Figure 6. The results show that when the utilization is low, the single-cell packet has the best performance. When the utilization is high, the SVP with size 2 has the best delay performance. This is because, when the link utilization is low, the packets are usually very sparse on the transmission link; hence, the packet transmission time through the switch dominates the delay performance. When the link utilization is high, the probability of continuous arrivals of packets with the same destination at the input port (called P_c) intensifies the output contention problem. When P_c is high, the average queue length is higher and each packet will experience a higher queueing delay. However, when the SVP size increases, P_c decreases. With a smaller P_c, the average queue length is reduced. Although the average queue length is reduced, the average queue delay is the product of the average queue length and the packet service time. Therefore, there is a tradeoff between the packet service time through the switch and the SVP size to optimize the cell delay under the same link utilization. In conclusion, the cell delay distribution through the switch is determined by two factors: the packet transmission delay and the probability of continuous arrivals of packets with the same destination at the input port.

From the cell delay distribution provided in Figure 6, using the extrapolation scheme the buffer size required to achieve CLR 10^{-9} is around 100 for single-cell transmission when the link utilization is only 0.55. To achieve the same CLR, the buffer size requirement will be increased exponentially when the link utilization approaches 0.6 switch throughput. Therefore, increasing the buffer size is not effective for improving the throughput of the fast packet switch, and it is not appropriate for the satellite environment.

3.2.1.2 Non-FIFO Queue

Using a non-FIFO queue with a checking depth of 2, the switch throughput can be increased from 0.6 to 0.73. Using a non-FIFO queue with a checking depth of 3, the switch throughput can be increased from 0.6 to 0.79. The switch throughput improvement decreases when the checking depth increases.

The cell delay jitter of the switch for checking depths of 2 and 3 is depicted in Figure 7. Compared this figure with Figure 4, the non-FIFO queue is an effective scheme for increasing the down-link beam utilization. In implementation, the maximum checking depth is determined by the processing speed of the routing tag of each packet at the input port queue. Since the scheduling algorithm must operate at the same speed as the switch and the number of packets processed by the scheduling algorithm is proportional to the switch size, the maximum checking depth allowed for a smaller switch is larger than that of a larger switch. Since the size of the on-board switch is between O(10) and O(100), the non-FIFO queue with a preset checking depth is proper for satellite application.

3.2.1.3 Increasing Switch Speed

The third scheme is to operate the switch at a speed faster than the link speed so that more incoming packets can be processed by the switch in one packet slot time (pkt size/link speed) and the output contention problem is reduced. If the output contention problem is reduced, the input queueing delay is also reduced. In this scheme, since the switch speed is greater than the down-link speed, to effectively improve the down-link beam utilization, buffering is required at the output ports to hold the packets. The output queue performs as a statistical multiplexer and the speed of the multiplexer is the same as the down-link speed.

It is possible to combine the non-FIFO queue scheme and the speedup scheme so that the tradeoff among throughput, delay, and hardware cost can be optimized. Two configurations simulated to show the effect of speedup: the first has a speedup factor of 1.5 and checking depth of 1, and the second has a speedup factor of 1.25 and checking depth of 2. Note that the maximum achievable throughput for both configurations is around 0.9.

The cell delay jitter of the switch for both configurations is depicted in Figure 8. The single-cell and SVP delay distributions for both configurations with link utilization as a parameter are provided in Figures 9 and 10, respectively. For single-cell transmission, when the utilization is less than 0.85, the delay performance for both

configurations is about the same. When the utilization approaches 0.9, the first configuration performs better. This is because the packet transmission time through the switch for a speedup factor of 1.5 is less the packet transmission time for a speedup factor of 1.25. For SVP transmission, the delay performance is about the same for both configurations and all utilizations. Note, as previously mentioned, that there is a tradeoff between the packet service time through the switch and SVP size to optimize the packet delay. Therefore, the combination scheme ,using speedup and a non-FIFO queue, is adequate for single-cell transmission and lower utilization, and it is appropriate for SVP transmission for all utilizations.

3.2.2 Output Buffering

In this approach, the packets are buffered at the output ports. To resolve output contention, either the switching fabric must operate at a speed faster than the line speed, as in the case of the banyan-type network, or there must be disjoint path between any input-output pair and the output port must have multiple buffers such as the knockout switch [3]. In this paper, only the banyan-type switch is considered. If the switching speed is N times faster than the line speed (where N is the size of the switch), all the packets destined to the same output port during the same slot time can be buffered at the output port. However, this approach is feasible only if the link speed is low and the switch size is small. In conclusion, it is not feasible to use output buffering alone to increase the throughput of the banyan-type switch.

4. Concluding Remarks

Based on the performance analysis of SVP transmission through the earth station, to fully utilize the SVP concept without affecting the delay quality, the up-link and down-link must operate at a very high utilization (above 80 percent). In such a configuration, the packet delay at the earth station is minimized. It is also found that the SVP formatting at the earth station reduces the probability of continuous arrivals of packets with the same destination at the input port. The result has shown to be effective in reducing output contention of the on-board switch. It suggests that spacing is one of the necessary requirements for satellite B-ISDN congestion control. The spacing process is one of the traffic shaping functions used to ensure that cell streams coming into the network do not exceed the negotiated value between the subscriber and the network. The spacing mechanism does not send the packets with the same destination back by back, since the peak value for this stream is the same as the satellite transmission link. Other packets with different destinations can be inserted between two packets with the same destination. The result of spacing is that output contention of the on-board switch is reduced.

To provide high up-link and down-link utilization within the satellite network, the on-board fast packet packet throughput must be greatly improved. Several schemes have been examined in the paper. Increase buffer size is not effective for the satellite environment. The non-FIFO queue with a preset checking depth at the input port to resolve HOL blocking has shown to be effective. However, the throughput improvement is still very limited since it is not practical to use a large checking depth. To significantly increase the throughput, the speedup scheme must be used. It has been found that the non-FIFO queue in conjunction with the speedup scheme can significantly increase the throughput of the switch at a reasonable hardware cost. Therefore, the combination of input queueing, a non-FIFO queue with a preset checking depth, speedup, and output queueing optimizes the performance of the switch and the hardware cost.

This paper has analyzed single-cell and SVP performance through the satellite B-ISDN for point-to-point traffic only. Extension of this work to point-to-multipoint traffic is currently under study.

References

[1]   F. A. Tobagi, "Fast Packet Switch Architectures for Broadband Integrated Services Digital Networks," Proc. IEEE, Vol. 78, No. 1, pp. 133–167, Jan. 1990.
[2]   M. Karol, M. Hluchyj, and S. Morgan, "Input vs Output Queueing on a Space-Division Packet Switch," IEEE Trans. on Communications, Vol. 35, pp. 1347-1356, Dec. 1987.
[3]   U. Yeh, M. Hluchyj, and A. Acampora, "The Knockout Switch: A Simple, Modular Architecture for High-Performance Packet Switching," *IEEE JSAC*, Vol. 5, pp. 1274–1283, Oct. 1987.
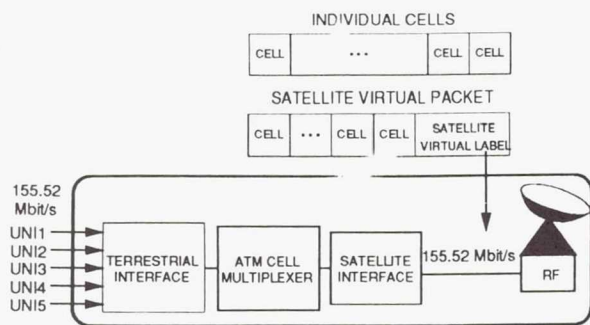
Figure 1: ATM Earth Station Configuration
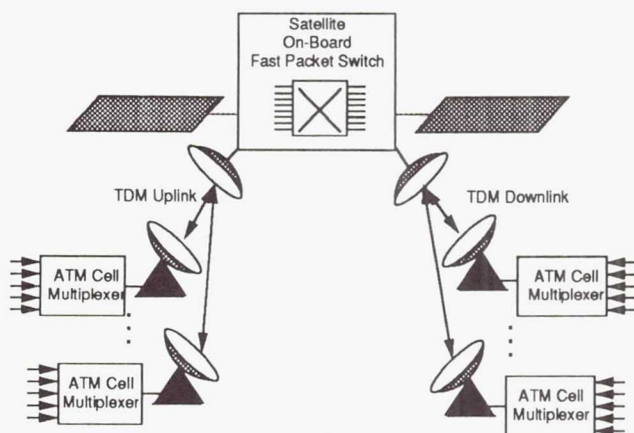
Figure 2: Satellie B-ISDN Simulation Model
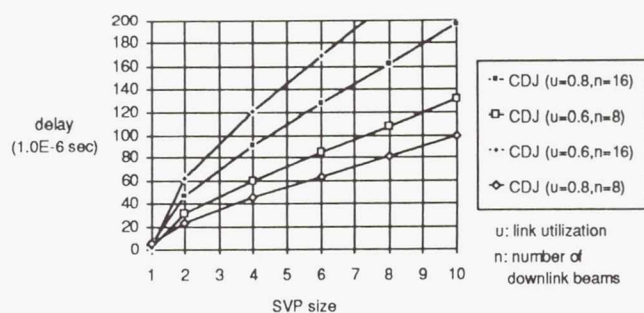
Figure 3: Cell Delay Jitter Versus SVP Sizes for Different
Link Utilization and Number of Downlink Beams
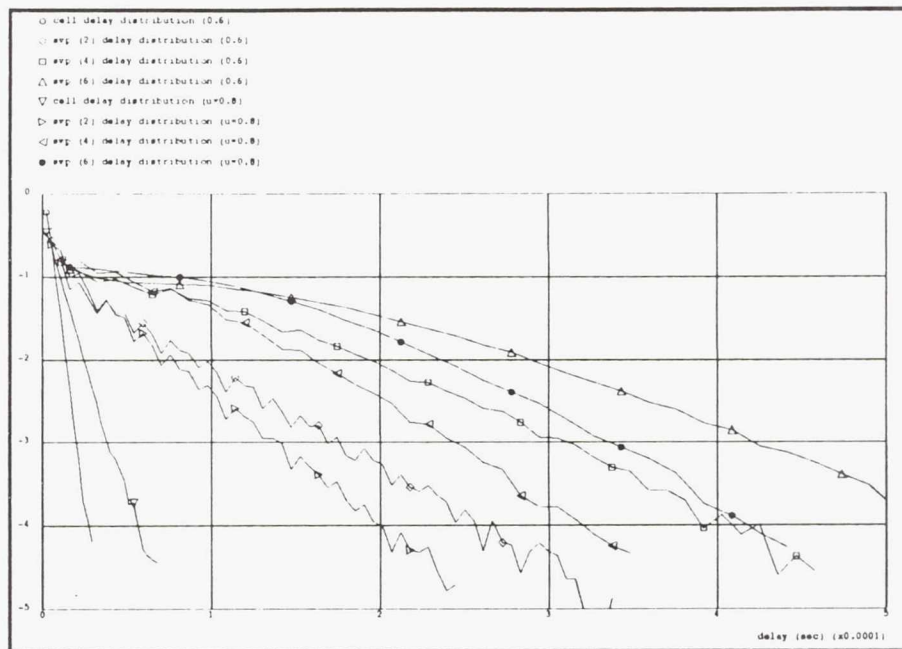
Figure 4: Cell Delay Distribution through the Earth Station
for Different SVP Sizes and Link Utilization

57
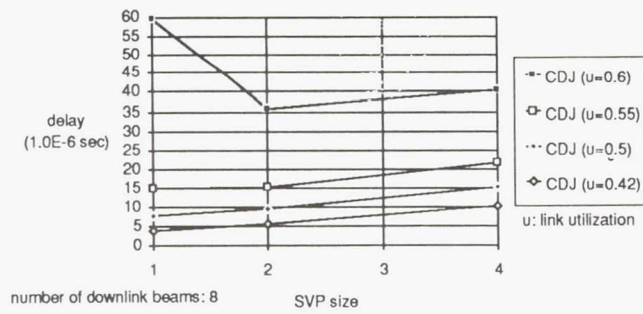
Figure 5: Cell Delay Jitter Versus SVP Sizes for Different
Link Utilization
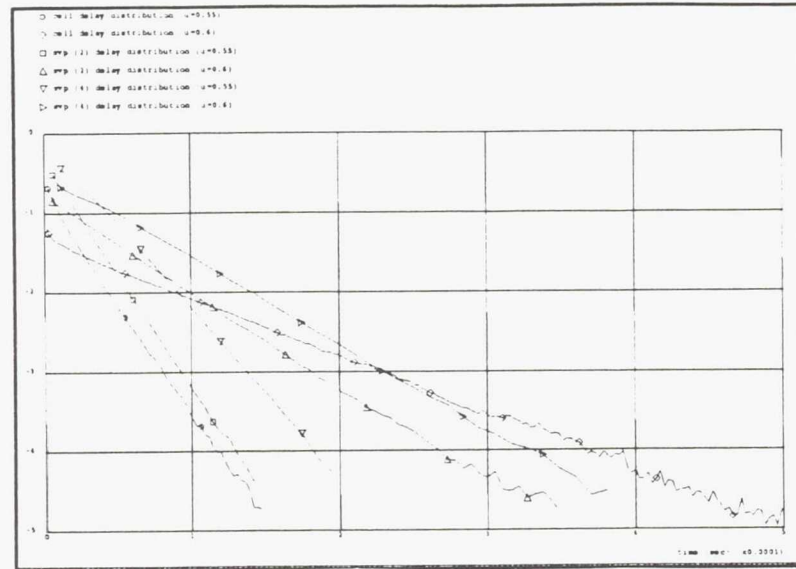
Figure 6: Cell Delay Distribution through the On-Board Switch for Different
SVP Sizes and Utilizations
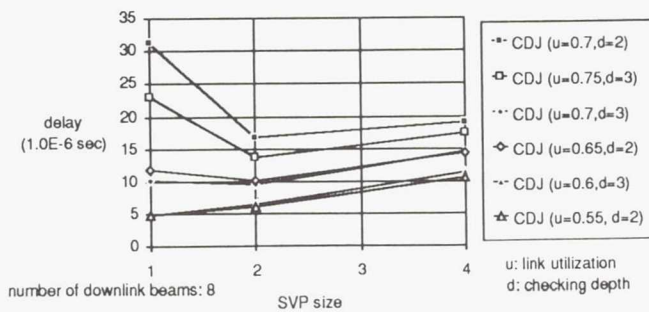
Figure 7: Cell Delay Jiiter Versus SVP Sizes for Different
Link Utilization and Checking Depth
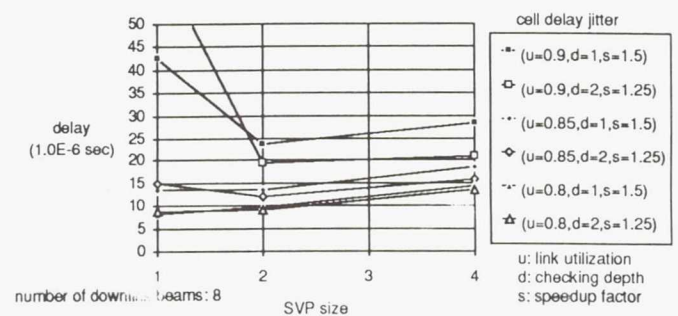
Figure 8: Cell Delay Jitter Versus SVP Sizes for Different
Link Utilization, Checking Depth, and Speedup Factors

58
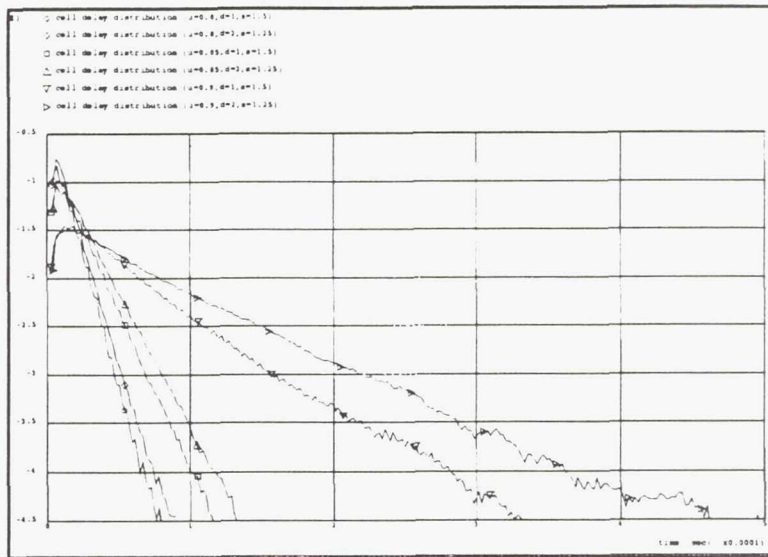
Figure 9: Single Cell Delay Distribution through the On-Board Switch for
Different Utilizations, Checking Depth, and Speedup Factors
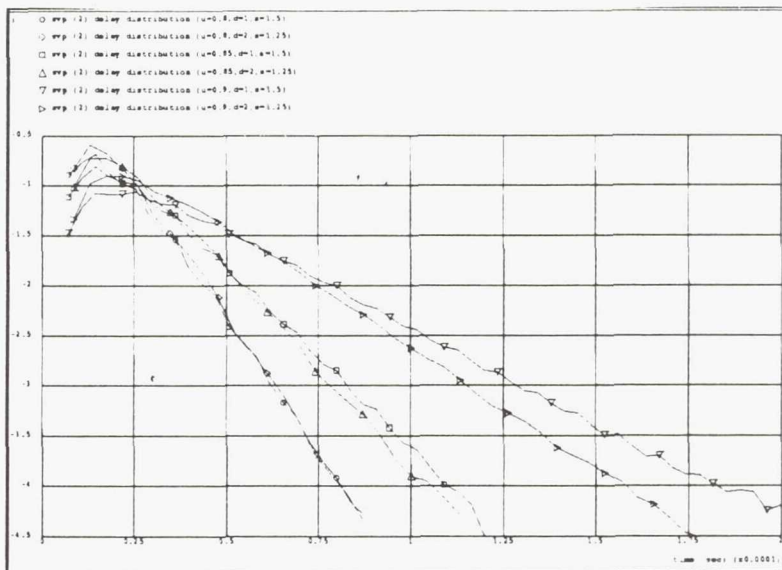
Figure 10: SVP Delay Distribution through the On-Board Switch for Different
Utilizations, Checking Depth, and Speedup Factors When SVP Size is 2

59

Page intentionally left blank

A MULTIDISCIPLINARY APPROACH TO THE DEVELOPMENT OF LOW-COST HIGH-PERFORMANCE LIGHTWAVE NETWORKS

Jacek Maitan and Alex Harwit
Lockheed Missiles & Space Company, Inc., Research & Development Division,
3251 Hanover Street, Palo Alto, CA 94304-1191
e-mail: jmaitan@isi.edu

ABSTRACT

Our research focuses on high-speed distributed systems. We anticipate that our results will allow the fabrication of low-cost networks employing multi-gigabit-per-second data links for space and military applications. The recent development of high-speed low-cost photonic components and new generations of microprocessors creates an opportunity to develop advanced large-scale distributed information systems. These systems currently involve hundreds of thousands of nodes and are made up of components and communications links that may fail during operation. In order to realize these systems, research is needed into technologies that foster adaptability and scaleability. Self-organizing mechanisms are needed to integrate a working fabric of large-scale distributed systems. The challenge is to fuse theory, technology, and development methodologies to construct a cost-effective, efficient, large-scale system.

SCOPE OF THE PROBLEM

Designers of future large-scale structures for space applications must be able to solve the problems associated with access to large amounts of data distributed at various sites. The large size and distributed nature of the space applications dictates that the data be accessed with low latency and wide bandwidth. Space distributed systems of the future will need to be more reliable and require less management than existing commercial networks. Once installed, the space network must be capable of detecting, diagnosing, and recovering from both software and hardware failures. The system must be maintained though the use of a highly decentralized control structure in order to accommodate rapid changes in the system configuration and traffic patterns. Thus, the reliability requirements favor a distributed control solution. The ultimate control objective is a high degree of continuous system availability [Maitan 89a, 89b]. The number of such large-scale distributed applications will continue to grow [Chlamtac 90, Hinton 88, Nussbaum 88].

This paper is organized into three parts. In the first part, we introduce the concept of a multigrid network architecture (MNA) [Maitan 90a, 90b, 91]. In the second, we discuss the operation of MNA and suggest possible insertion points for new photonic technologies. Finally, we discuss approaches to increase the performance of MNA from 1 Gbps using state-of-the-art electronics to 50 Gbps per link using anticipated lightwave technologies. In this discussion we focus on issues associated with combining high-speed networks with low-overhead protocols and how this process affects the architecture.

ISSUES

Data networks for computer communication must offer high bandwidth (gigabits-per-second), and low latency [Young 87] and be able to handle highly variable multimedia traffic [Lidinsky 90]. Existing solutions such as high-performance parallel interface (HIPPI) or asynchronous transfer mode (ATM) were

designed to address some of these needs. HIPPI is designed to handle point-to-point data transfers only. Although ATM implementations are fast, ATM requires the path to be established before transferring data. The approach works well for telephony; however, it is insufficient when a single computer broadcasts data to a large group of computers. We conclude, therefore, that systems based on the existing standards may not be able to offer effective solutions to space networking needs.

Fiber-optic networks are characterized by a small ratio between the packet transmission time and the packet propagation time through the network. The latter is usually larger because it includes the time required to route and to resolve contention at nodes. In fiber-optic networks, this is the major source of propagation delays. An increase in the transceiver speed to decrease the packet transmission time will not increase the actual network throughput.

Increased transmission speeds change the data communication strategy. Bandwidth is cheap and computation is expensive. In traditional connection-oriented systems, the traffic is controlled by a simple store-and-forward algorithm. This strategy requires careful management of the buffering and protocol processes [Clark 89], and is effective only when the computers are much faster than the networks. In state-of-the-art systems, i.e., a 32-bit computer operating at 100 million instructions per second (MIPS), the total data flow is 3.2 Gbps. This is the same order of magnitude as the 1 Gbps available in a communication link. Thus, the ability to handle high-speed traffic by simple store-and-forward algorithms using general purpose computers has disappeared.

The key solution to all of these problems is an effective control strategy. Today's careful bit stuffing, used to increase bandwidth utilization, results in complex protocols, which must be replaced with bandwidth-effective data transmission protocols with lower processing overhead.

MULTIGRID NETWORK ARCHITECTURE

Our research has focused on providing a network which is flexible, scaleable, simple to control, and very effective in its operation. To address these objectives, we simplify the medium access control (MAC) protocol and eliminate as much of the protocol processing overhead as possible. This begins with the elimination of the store-and-forward transport algorithms. The result is a novel bufferless control structure for very high-performance packet-switching networks. A full multigrid network architecture (MNA) implementation features both circuit and packet switching [Maitan 90c]. In this paper, the discussion is limited to packet switching only.

A network switch, also referred to as a node, is shown in Fig. 1. Here, a data packet must bid for an output link by providing an ordered list of links which can be used. The router must simply assign an output link for each packet, based on all the bids. Conflicting bids for the same output port are nondeterministically solved. Thus, the whole network is treated as both a buffer and a distributed data processing structure that is capable of routing packets toward their desired destination. The bidding process is strictly local and solves the problem of central control, which would otherwise be nearly impossible to achieve. MNA is a connectionless network and packets from all routing nodes and hosts are treated equally when competing for resources. No restriction is placed on the topology or size of the network.

Most existing networks use dynamic routing, in which priorities are computed based on traffic estimates. Thus, if not controlled, the asynchronous nature of dynamic routing table updates may result in packet looping. In MNA, the tables are statically computed and permanent loops can be detected and eliminated. In an MNA routing table, a failed link is simply marked as a busy link. This routing mechanism substantially simplifies the complexity of controlling the node and, as we have demonstrated, leads to a very simple hardware system implementation.
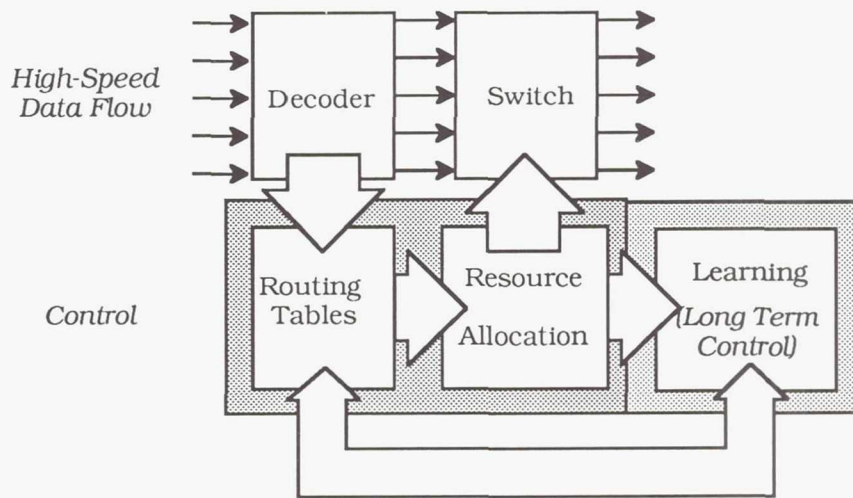
Fig. 1 A simplified block diagram of a router for a packet switch

In a well-connected network of arbitrary topology, there is usually more than one output link suitable for routing, Fig. 2. Access to a particular link is given using a stochastic resource-allocation policy and priority is given to packets that are already in traffic. Thus, in highly congested traffic, extra packet bursts diffuse to neighboring nodes, so that, the network as a whole is able to store extra traffic. The failure of a link may cause an imbalance in the system. For example, if the number of incoming packets at a node exceeds the number of outgoing links, packets may be lost. Higher level protocols must be used to control the recovery of missing data as discussed later in this paper.
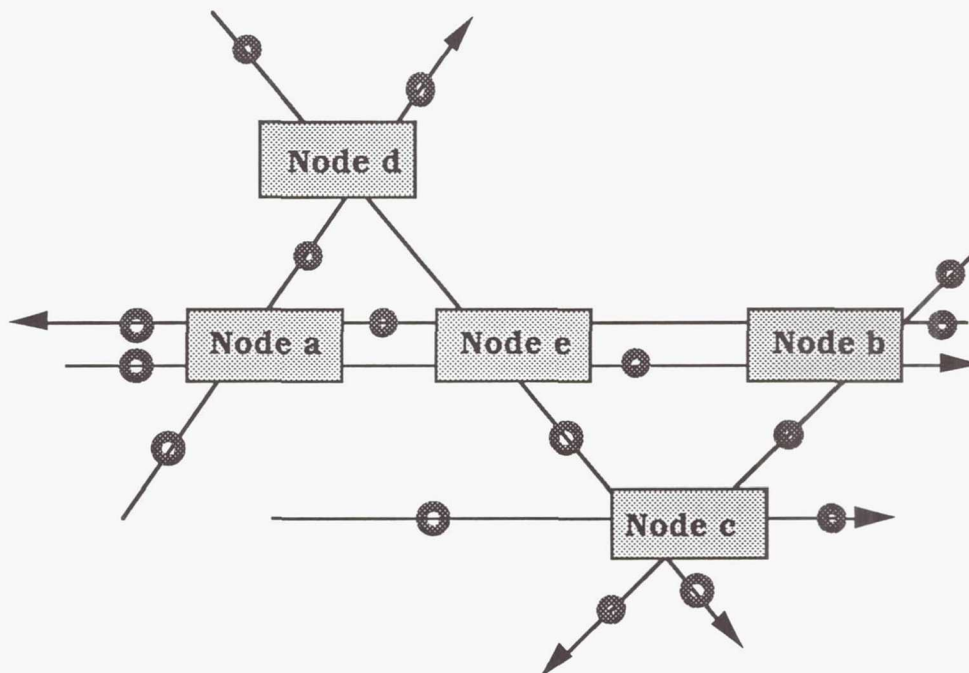
Fig. 2 Interaction between nodes in MNA

63

We have built prototype hardware and performed computer simulations [Gburzynski 90] to show that a distributed control scheme of this type applied to such networks is very effective for a wide range of parameters. The lengths of the routing paths are on average independent of load and the physical lengths of the links.

MNA PROTOCOL

All data is transferred as fixed-size packets consisting of a header and a payload. The header contains control and error-detection information in addition to the destination address. Packets can be sent as single entities or in groups. They are generated by the hosts which are connected to the routing nodes. To be transferred, a packet must first gain access to the transport fabric.

On arrival at a node, a packet submits a bid to the resource allocator for an output port. If all desired ports are busy, an arbitrary free port is assigned to keep a packet in transition. A packet is lost only when no free output ports are available. Once access to a port is granted, it is used to transmit the entire packet. MNA does not use local storage; instead, it uses the whole network as a buffer.

In the case of substantial system failures combined with a heavy load, packet losses are unavoidable and must be handled by an appropriate transport protocol. Thus, under catastrophic conditions, MNA converges to the traditional connection-oriented network with circuit reservation. However, unlike traditional fault-tolerant networks, all resources are used; none are kept simply as a reserve. The soft failure feature is built into MNA and is not added as an afterthought.

In all-optical networks, one can split the signal and process only the header. This lack of intermediate buffering simplifies the construction of an all-optical switching network, as discussed below.

MNA SYSTEM INTEGRATION

In this section, we outline the approach we are taking in the hardware prototype that is currently under construction. We also discuss how MNA scales up with new high-speed technologies.

The packet-switching circuitry is currently being constructed and tested in an all-electrical implementation. It is composed of off-the-shelf CMOS devices and gate arrays. To date, we have constructed and tested circuits to process an estimated 10^7 resource allocation bids per second. In the future, the network is envisioned to utilize an all optical switching fabric.

In a photonic network, the nodes are connected to each other by optical fibers. Figure 3 shows a block diagram of an 8 x 8 high-speed network switch. This switch consists of two 8-input fiber couplers and routing circuitry to route incoming data to the appropriate output. Data packets are assumed to arrive on single-mode 1550-nm optical fibers, formatted in 500-bit packets at 50 Gbps. The total packet length is 10 ns.

Optical data enters the switch through the input fiber coupler. In order to correctly route an incoming packet, a small amount of optical energy is removed from the incoming signal to form the control path. Energy in the control path is converted to an electrical signal from which a destination address is extracted. The destination address is then used as the input for a routing circuit that performs resource allocation and finally configures the switch. The remainder of the signal flows into a temporary buffer which consists of a small length of Er^+-doped fiber that acts both as a delay line and wide-bandwidth amplifier to boost the optical signal. After amplification, the output of the buffer goes to an optical switch which routes the data to the appropriate output on the output fiber coupler.
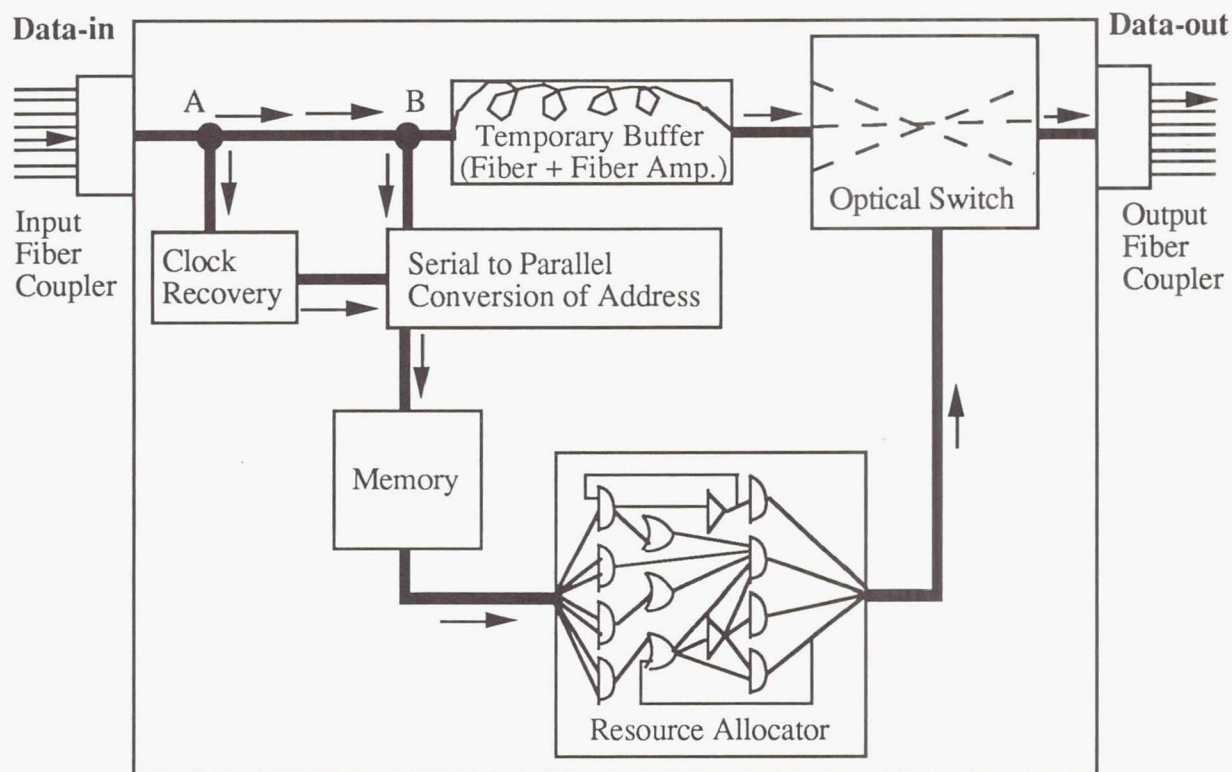
Fig. 3   An 8 x 8 high speed network switch

In the control path, the switch must recover the clock and extract the header of the packet. Specifically, a small amount of optical energy is tapped off at point "A" for a clock-recovery circuit [Swartz, 88]. The function of this circuit is to align the clock of the node that generated the data with the local clock of the switch. There is one clock-recovery circuit per input line. At point "B", a second optical tap removes additional signal and sends it to a serial-to-parallel converter. With the clock timing information from the clock-recovery circuit, this circuit extracts the destination address from the packet header and converts it to a parallel format. The destination address is then used as an address for a routing table that is simply a high-speed memory. Each memory location is associated with a destination and contains an ordered list of desired output channels to be used as bids. There is one serial-to-parallel converter and one memory for each input line. The bids from each memory are all processed by a single resource-allocator unit. The resource-allocator unit is composed of a large sequence of discrete logic gates. The output of the resource allocator configures an optical switch. To properly control the switch, the spacing between packets must be greater than the sum of the optical switch switching and settling times.

Several constraints exist on each of the circuits in the network switch. The fiber coupler, for example, must be able to handle at least eight single-mode 1550-nm fibers with a low insertion loss. It will most likely be composed of a V-groove technology in which the fibers are pigtailed into the substrate containing the electronic processing circuitry. Since each packet is about 10 ns long, the fiber line should delay the signal by about 20 ns, during which time the circuit could process the packet destination address and set the optical switch. The clock extraction circuit must be able to extract the 50 Gbps timing in about 3 ns. The serial-to-parallel converter must also be able to extract the packet destination address in about 3 ns. For an 8 x 8 switch utilizing a bid format consisting of three options, the memory size would be equal to: [number of nodes in network] x [3 x $\log_2$(number of nodes)]. To balance the flow in the data path, the memory is required to have an access time of about 4 ns, and such parts are available today. The resource allocator is composed of discrete "AND", "OR", and "NOR" gates, etc., and would be required to switch in about 4 ns. The low-loss optical switch may either be an integrated optoelectronic polymer switch with

active rail taps [Van Eck 91] or a multiple-quantum-well modulator [Komatsu, 90]. It is required to switch in about 4 ns.

The key point in the MNA approach is the reduction of the packet propagation time by careful management of routing information combined with the application of a new class of evaluators to resolve synchronization and resource-allocation problems. This is especially important in local area networks (LANs) where one must also be able to manage latency. Instead of optimizing protocol layers one at the time, we have attempted to consider the interaction of mechanisms associated with several protocol layers simultaneously.

SUMMARY

In this paper, we have discussed a completely distributed packet-routing architecture called multigrid network architecture (MNA), for building low-cost high-speed networks. To prove the feasibility of such networks, we have prototyped a resource allocator which has an estimated performance of 8×10^6 packets/s for an 8 x 8 packet switch. Currently, we are completing a hardware prototype of an 8 x 8 pizza-box-sized packet switch capable of handling 1 Gbps traffic at each port. Furthermore, the networks are capable of transferring data at up to 50 Gbps per line and can be controlled using MNA distributed-control algorithms.

MNA is an architecture that has been designed to scale with evolving technologies. It is also an attempt to simplify and integrate an implementation of a multilayer protocol stack. This work is an attempt to identify an approach leading to the cost-effective use of high-speed networks in application-oriented distributed systems. Preliminary results are encouraging and indicate that such networks can be controlled using simple algorithms implemented in low-volume switches that can be built using existing technologies.

ACKNOWLEDGMENTS

REFERENCES

[Chlamtac 90]   Chlamtac, I., and Franta, W.R., "Rationale, Directions, and Issues Surrounding High-Speed Networks," *IEEE Proceedings,* Vol. 78, No. 1, pp. 94-120, January 1990.

[Clark 89]   Clark, D.D., van Jacobson, D., Romkey, J., and Salwen, H., "An Analysis of TCP Processing Overhead," *IEEE Communications Magazine,* Vol. 27, No. 6, pp. 23-36, June, 1989.

[Gburzynski 90]   Gburzynski, P., and Rudnicki, P., *The LANSF Protocol Modeling Environment, version 2.0,* Department of Computer Science, University of Alberta, Edmonton, Alberta, Canada, 1990.

[Hinton 88]   H. S. Hinton. "Architectural Considerations for Photonic Switching Networks". *IEEE Journal on Selected Areas in Communications,* Vol. 6, No. 7, pp. 1209-1226 August 1988.

[Komatsu, 90]   Komatsu, K., Sugimoto, M., Ajisawa, A., Hamamoto, K., Kohga, Y., and Suzuki, A., "4 X 4 GaAs/AlGaAs Optical Matrix Switches with Electro-Optic Guided-Wave Directional Couplers", in *Photonic Switching II*, editors Tada, K and Hinton, H. S. , Springer Verlag, Berlin, 67-71, 1990.

[Lidinsky 90]	Lidinsky, W. P., Data Communications Needs, *IEEE Network,* Vol. 2, No. 4, pp.28-33, March 1990.
[Maitan 89a]	Maitan, J., "Intelligent Distributed and Networking Systems", *Proceedings of IJCAI workshop on The Future of Research in Artificial Intelligence,* IFIP's Artificial Intelligence Specialists Group sponsored by IFIP and IJCAI, 1989.
[Maitan 89b]	Maitan, J., Ras, Z., and Zemankova, M., "Querry Handling and Learning in a Distributed Intelligent System", *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems,* Charlotte, North Carolina, 1989.
[Maitan 90a]	Maitan, J., Walichiewicz, L., and Wealand, B., "Communication Technology Requirements for the Next Generation of Ground Systems," Final Report, May 1990.
[Maitan 90b]	Maitan, J., Ras, Z.W., "OPERA: Opto-Electronic Reconfigurable Architecture for Multiprocessor Systems, Mathematical Foundations", *Proceedings of the Fifth International Symposium on Methodologies for Intelligent Systems,* Charlotte, North Carolina, 1990.
[Maitan 90c]	Maitan, J., Walichiewicz, L., and Wealand, B., "Integrated Communication and Information Fabric for Space Applications", *AIAA/NASA Second International Symposium on Space Information Systems,* September, 17-19, 1990
[Maitan 91]	Maitan, J.: "A Flow Control Mechanism for Distributed Systems," *Proceedings of the SPIE Conference on Target Recognition,* Orlando Florida, 1991.
[Nussbaum 88]	Nussbaum, E. "Communication Networks Needs and Technologies - A Place for Photonic Switch?" *IEEE Journal on Selected Areas in Communications,* Vol. 6, No. 7, pp. 1036-1044, August 1988.
[Swartz, 88]	Swartz, Robert G., "High Performance Integrated Circuits for Lightwave Systems", in *Optical Fiber Telecommunications II,* editors Miller, Stewarte and Kaminow, Ivan P., Academic Press, Boston, 759-780, 1988.
[Van Eck 91]	Van Eck, T. E., Ticknor, A. J. , Lytel, R. S. , and Lipscomb, G. F., "Complementary Optical Tap Fabricated in an Electro-optic Polymer Waveguide", *Appl. Phys. Lett.* Vol 58, pp. 1588-1590, 1991.
[Young 87]	Young, M., Tevanian, A., Rashid, R., Golub, D., *The Duality of Memory and Communication in the Implementation of a Multiprocessor Operating System,* unpublished material, Department of Computer Science, Carnegie-Mellon University, 1987.

Page intentionally left blank

Architecture for Survivable System Processing (ASSP)

Richard J. Wood
Rome Air Development Center (RADC/OCTS)
Griffiss AFB NY 13441-5700
(315) 330-2641

SUMMARY:

The ASSP Program is a multi-phase effort to implement DOD and commercially developed high-tech hardware, software and architectures for reliable space avionics and ground based systems. System configuration options provide processing capabilities to address Time Dependent Processing (TDP), Object Dependent Processing (ODP) and Mission Dependent Processing (MDP) requirements through Open System Architecture (OSA) alternatives that allow for the enhancements, incorporation and capitalization of a broad range of development assets. High technology developments of hardware, software, networking models address technology challenges of long processor life times, fault tolerance, reliability, throughput, memories, radiation hardening, size, weight, power (SWAP) and security.

Hardware and software design, development and implementations focus on the interconnectivity/interoperability of an open system architecture and is being developed to apply new technology into practical OSA components. To insure for widely acceptable architecture capable of interfacing with various commercial/military components, this Program provides for regular interactions with Standardization Working groups (eg) the International Standards Organization (ISO), American National Standards Institute (ANSI), Society of Automotive Engineers (SAE), and Institute of Electrical and Electronic Engineer (IEEE). Selection of a viable open architecture is based on the widely accepted standards that implement the ISO/OSI Reference Model.

DEVELOPMENT:

The ASSP Program provides research and development tasks to implement heterogeneous processing nodes of various configurations into the OSA network. Each node resides on a single circuit card with onboard scalar, vector processing components. (See Figure 1).
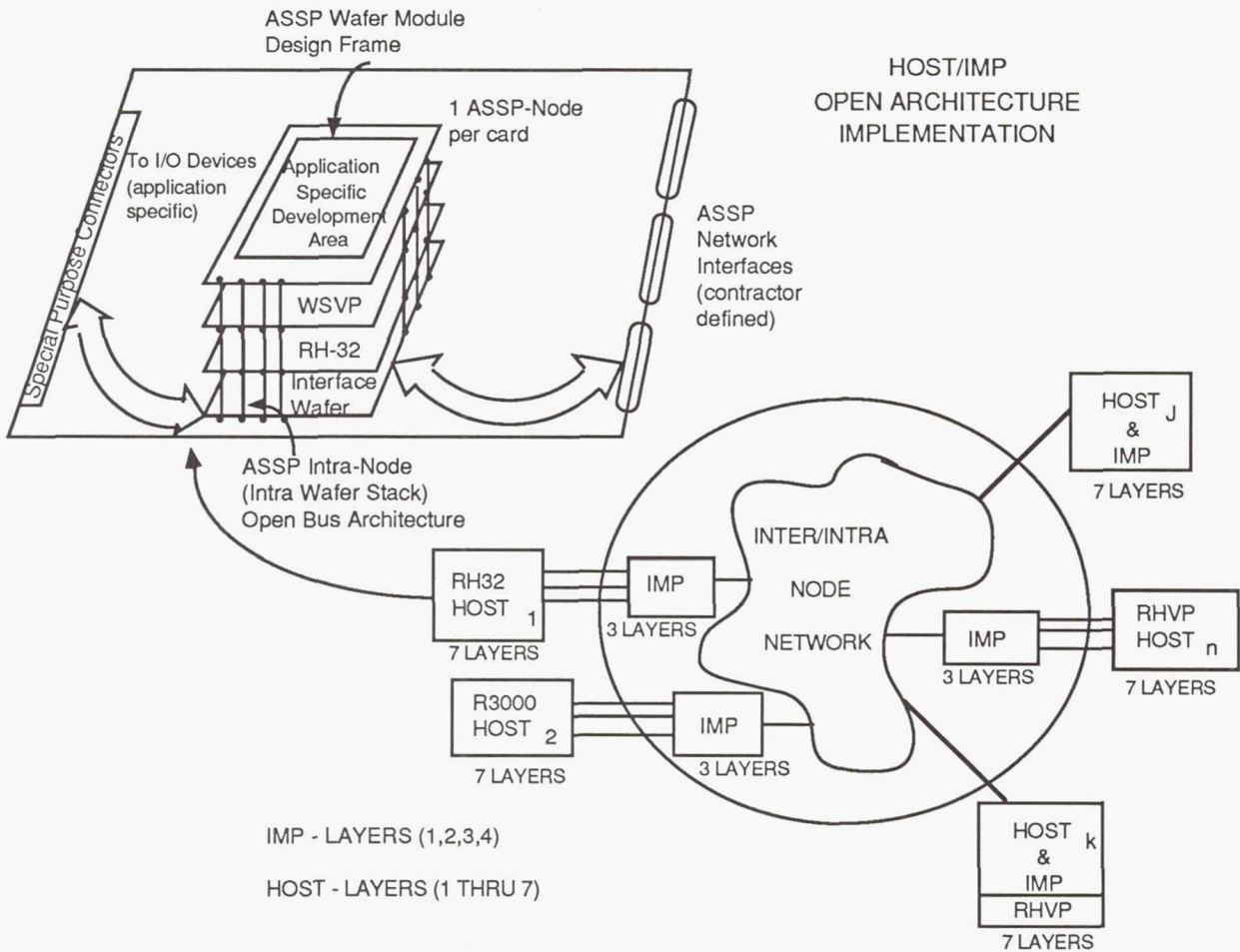
FIGURE 1

Simulation and demonstrations will be accomplished to incorporate processing components (i.e. 3D Computer, AOSP LAN, large memories, associated co-processor) into the expandable open architecture. As such, key features include standard bus hardware/protocols, support for tightly coupled and/or loosely coupled multiprocessing, and object oriented operating system primitives. The ASSP will build on the successes of the Advanced Onboard Signal Processor (AOSP) program by furthering system reliability through shrinking node size, improving reliability at each level of the architecture through incorporation of fault tolerance techniques, and exploiting the latest advances in fault tolerant, secure operating systems design. As hardware development technology provides more capable and innovative components, such as radiation hardened, WSI, HDI, Photonics, Wafer Stacking and application specific integrated circuits (ASIC), they will be implemented into the ASSP phase two effort. Integration along with developments of form fit factors to support physical space budgets will be demonstrated as an Advanced Development Model (ADM).

The major thrust of this program will address the key technical challenges for:

a. Open system architectures for rapid insertion of commercial/military technology
b. Interoperability/Interchangeability of heterogeneous processing nodes
c. Architectural incorporation of stacked hybrid wafer integration
d. Fault tolerant, real-time ADA Run-time systems for distributed heterogeneous processors

The ASSP program also includes implementation of industry and/or military hardware/software in accordance with standards that conform to the International Standards Organization, Open System Interconnection (ISO/OSI) Reference model (ISO 7498). To meet these objectives, advanced state of the art language and modeling designs/developments will be made to adequately represent each of the ISO/OSI levels, their interactions, and additions. (See Figure 2).
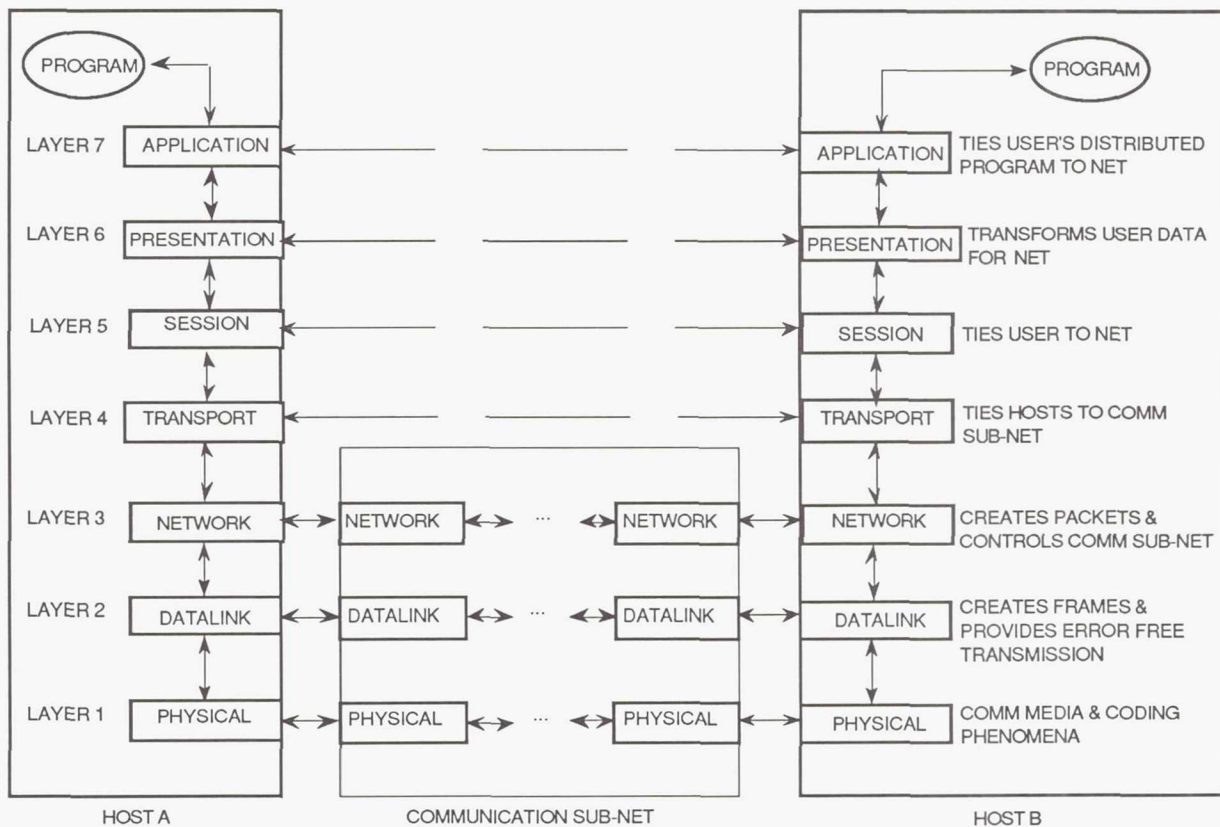
ISO/OSI REFERENCE MODEL

FIGURE 2

The Reference model will provide specifications for networks, backplanes, interfaces, and busses to support open systems. Model simulations will be developed and tested to ensure conformance with objectives that provides commonality, high performance, availability, and fault tolerance.

Implementation of current high technology components in this effort such as the WSVP, RH32, RHVP, WSI-HDI, Radiation Hardened Memories as well as commercial components like MIPS R3000 and Intel i860 processors in an "open architecture" processor network will be seriously considered for integration and demonstrations.

The two phased program is a five year effort with the first phase being two years duration. During this first phase, intensive studies of commercial and military hardware/software systems and components will be made to assess applicability for integration into the OSA. The design of an architecture adhering to accepted standards and responsive to space based applications will then be performed. The selected design will consider Intra-Nodal multiprocessing networks (Sub-net, or within a node or single board), Inter-Nodal Multi-processor networks (array of processors loosely or tightly coupled) and Inter-Satellite networks (Super-net, loosely coupled, communications may be micro-wave). (See Figure 3).

NODE WOTHIN NODE

SATELLITE NODES

Flexibility

Reliability

F.T.
Secure
ISO/OSI Model
Operating System

Inter-Communications

• STRESS MICROWAVE
LASER COMMUNICATION

• ARCHITECTURE
INDEPENDENCY

ADM NETWORK

• NETWORK - SIMULATOR
VALIDATION
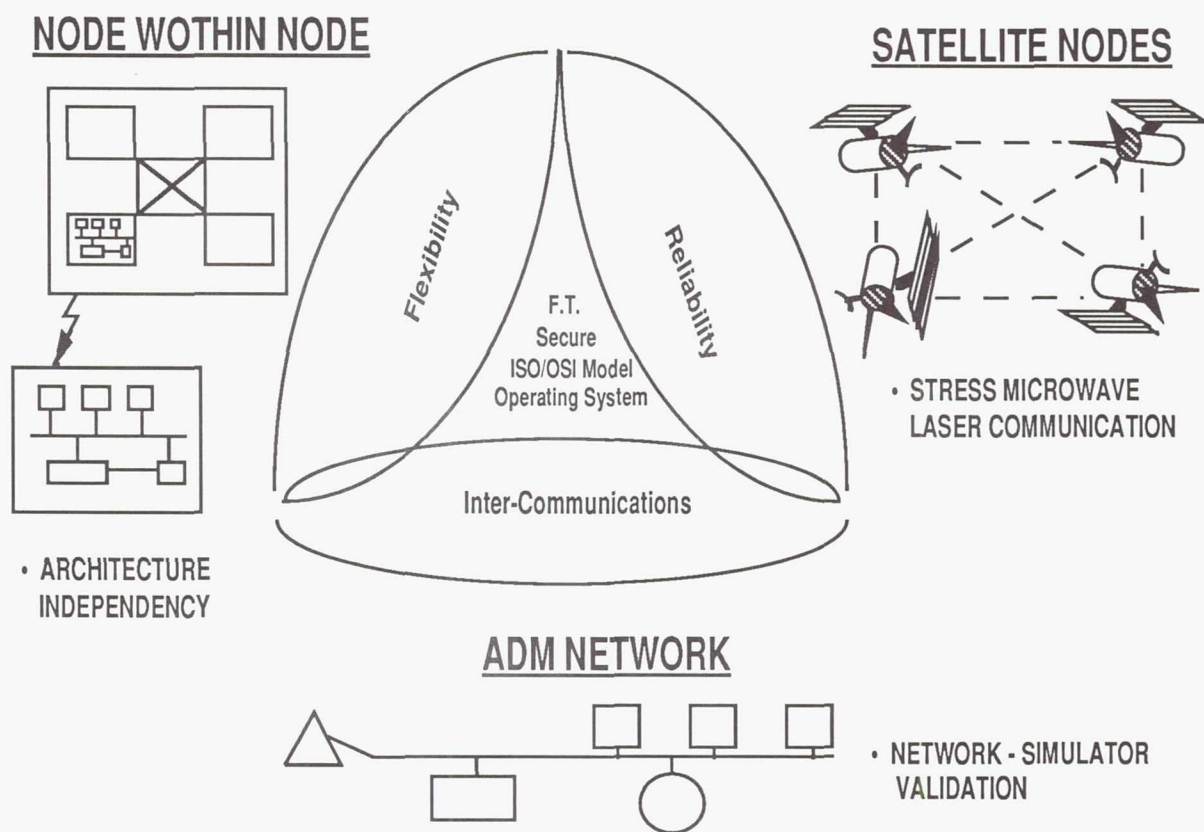
FIGURE 3

During the study and evaluation period, research in network simulators and tools will also be accomplished. New developments will be undertaken to support requirements where deficiencies occur. A final simulator configuration will then be selected and developed that will represent the Inter-Nodal Multi-Processor Network (INMPN) and also be the baseline design to support further development of the simulators for the Subnet and Supernet configurations. The INMPN simulator will be

used to prove the breadboard design concept of the OSA selected prior to the Preliminary Design Review (PDR).

The Phase I activities develop technology to control the open system via operating software that conform to layered protocols that implement the ISO/OSI model. Investigation of capabilities and potential applications of POSIX, SAFENET, NOS/GOS, and GOSIP will be accomplished, as well as hardware components such as Wafer Scale Vector Processors, military and commercial RISC/CISC processors. These components will be "retro-fitted" into the OSA via bus interface units (BIU's) to prove feasibility and acceptability of heterogeneous processors to operate as an integrated system.

Successful completion of Phase I will provide an operational application demonstration with a breadboard model of open system design exercising the heterogeneity of hosts. Capabilities will include graceful degradation, error recovery, dynamic routing and high performance capabilities with minimal latencies. The breadboard model along with respective simulations will provide the basis for implementation validation and design verification. Specifications, simulators, software development platform, and the baseline Open System Architecture will then be transitioned to Phase II.

Phase II is a three year program which basically reduces into hardware design the results obtained from Phase I. However, because of the updating of standards and to take advantage of new technology the Phase I architectural design will be refined to maximize responsiveness to the user community.

The Industrial/Commercial community has already widely accepted the standardization processes being offered by the ISO, ANSI, IEE etc. Therefore, Phase II will take advantage of this cooperation and inturn incorporate commercial breakthroughs in processor, communications and network technologies. Radiation hardened components, high technology processors such as the 3D computer, Gallium Arsenide developments, Photonics and Opto-Electronics technologies will be incorporated into the Open Systems where applicable to further advance the state of the art in network processing OSA. Components will be integrated without BIU's as the standardization regimens will dictate requirements to meet interoperability/interchangeability criteria.

To insure the proper application of developed standards (the ASSP program will NOT develop standards) associated contractors and the Technical Advisory Group (TAG) will work very closely with Standardization Working Groups of the various standardization communities. The TAG will be composed of government only experts in the fields of networking, fault-tolerance, security, reliability-maintainability, signal-data processing, memories, software and packaging. The TAG will insure that proper design and architectural plans are acceptable and representative of the government's interest. That the Architecture selected is widely acceptable by the military, industrial, commercial complex and where ever possible insure that potential standards specifically attributable to space based applications are considered by the Standardization committees.

Successful completion of Phase II will provide an Advanced Development Model (ADM) that will demonstrate interoperability/interchangeability along with the

above iterated assets in an Open Systems Architecture Network. This will provide the acceptable standard multiprocessor interconnects, high performance backplanes, switch networks, and network operating system.

The ASSP directly responds to an AIR FORCE, SDIO deficiency in providing an architecture that can support and upgrade processing systems without major redesigns, and procurements. This program also provides capabilities to launch processing networks that are versatile, offer various levels of complexity and are capable of rapid upgrades in mission profiles, hardware, and operating systems. The capability to incorporate commercial hardware breakthroughs along with their respective software support in a very short time frame and with a minimum of redesign/retooling is most beneficial and advantageous to the military-commercial community.

A Bandwidth Efficient Coding Scheme for the Hubble Space Telescope[*]

Steven S. Pietrobon
University of South Australia
The Levels, South Australia 5095
Australia

Daniel J. Costello, Jr.
University of Notre Dame
Notre Dame, Indiana 46556
U.S.A.

1 SUMMARY

As a demonstration of the performance capabilities of trellis codes using multidimensional signal sets, a Viterbi decoder for one of the codes in [1] was designed. The choice of code was based on two factors.

The first factor was its application as a possible replacement for the coding scheme currently used on the Hubble Space Telescope (HST). The HST at present uses the rate 1/3 $v = 6$ (with $2^v = 64$ states) convolutional code with BPSK modulation. With the modulator restricted to 3 Msym/s, this implies a data rate of only 1 Mbit/s, since the bandwidth efficiency $K = 1/3$ bit/sym. This is a very bandwidth inefficient scheme, although the system has the advantage of simplicity and large coding gain.

The basic requirement from NASA was for a scheme that has as large a K as possible. Since a satellite channel was being used, 8PSK modulation was selected. This allows a K of between 2 and 3 bit/sym. The next influencing factor was INTELSAT's intention of transmitting the SONET 155.52 Mbit/s standard data rate over the 72 MHz transponders on its satellites. This requires a bandwidth efficiency of around 2.5 bit/sym. A

Reed-Solomon block code is used as an outer code to give very low bit error rates (BER).

The 16 state rate 5/6, 2.5 bit/sym, 4D-8PSK trellis code from [1] was selected. This code has reasonable complexity and has a coding gain of 4.8 dB compared to uncoded 8PSK [2]. This trellis code also has the advantage that it is 45° rotationally invariant. This means that the decoder needs only to synchronise to one of the two naturally mapped 8PSK signals in the signal set.

2 ENCODER IMPLEMENTATION

At first, a systematic encoder was used in the design. However, it was found that in designing a Viterbi decoder, it would be simpler if a non-systematic convolutional encoder was used. This is because the state transitions in a non-systematic encoder are highly structured, compared with the almost "random" transitions of a systematic encoder.

To convert the systematic encoder to a non-systematic form, the technique described in [3] is used. This method uses the fact that the impulse response of each shift register in a non-systematic encoder will produce output sequences that are equivalent to the generator polynomials. Since a systematic encoder must also produce the same sequences, it is relatively easy to find k linearly independent output sequences from a systematic encoder that

Non-systematic Convolutional Encoder

Differential Encoder
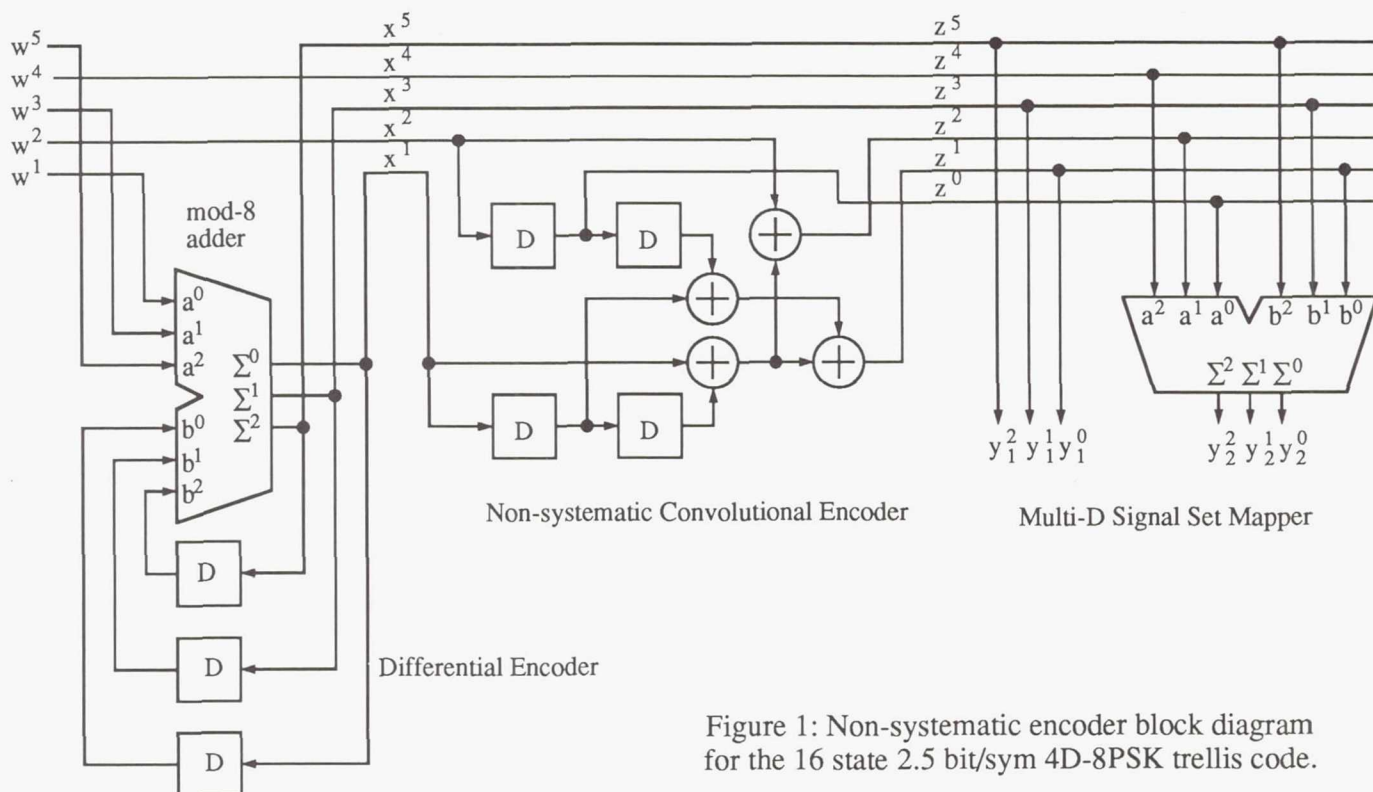
Multi-D Signal Set Mapper

Figure 1: Non-systematic encoder block diagram for the 16 state 2.5 bit/sym 4D-8PSK trellis code.

can be used as generators of a non-systematic encoder.

There is usually more than one set of possible generator polynomials. The polynomials are chosen so that the inputs $x^1(D)$ and $x^2(D)$ are affected by a 45° phase rotation in the same way as in a systematic encoder. Thus, the differential encoder for the systematic code can also be used for the non-systematic encoder. The non-systematic encoder equations that were found for the 4D-8PSK code are

$$z^2(D) = x^2(D) \oplus (D^2 \oplus 1)x^1(D), \qquad (1a)$$

$$z^1(D) = D^2x^2(D) \oplus (D^2 \oplus D \oplus 1)x^1(D), \qquad (1b)$$

$$z^0(D) = Dx^2(D). \qquad (1c)$$

Figure 1 illustrates the new non-systematic encoder. After a 45° phase rotation, we have $z^2_r(D) = z^2(D)$, $z^1_r(D) = z^1(D) \oplus 1(D)$, and $z^0_r(D) = z^0(D)$. Rotating the equations in (1) gives $x^2_r(D) = x^2(D)$ and $x^1_r(D) = x^1(D) \oplus 1(D)$, the same as for the systematic encoder.

The encoder uses a Phase Locked Loop (PLL) to generate the two times clock for transmitting the two 2D symbols. This PLL is based on the 74HC4046 Integrated Circuit (IC). The encoder is able to accept data either serially or in five bit bytes.

3 DECODER IMPLEMENTATION

Due to the complexity of the decoder design, only a brief description is given here. As such, only the important design decisions are described.

To reduce the cost of the codec, a serial implementation of the decoder was chosen. That is, one clock cycle would be required for each state of the code. Since there are 16 states, at least 16 clock cycles are required to process each received 4D point. As will be described in more detail later, an extra seven clock cycles are required for start-up purposes. Thus, a total of 23 clock cycles are required for each iteration of the Viterbi algorithm.

The technology and clock speed in our design is the same as used in another Viterbi decoder designed by the author [4]. This gave us greater confidence that
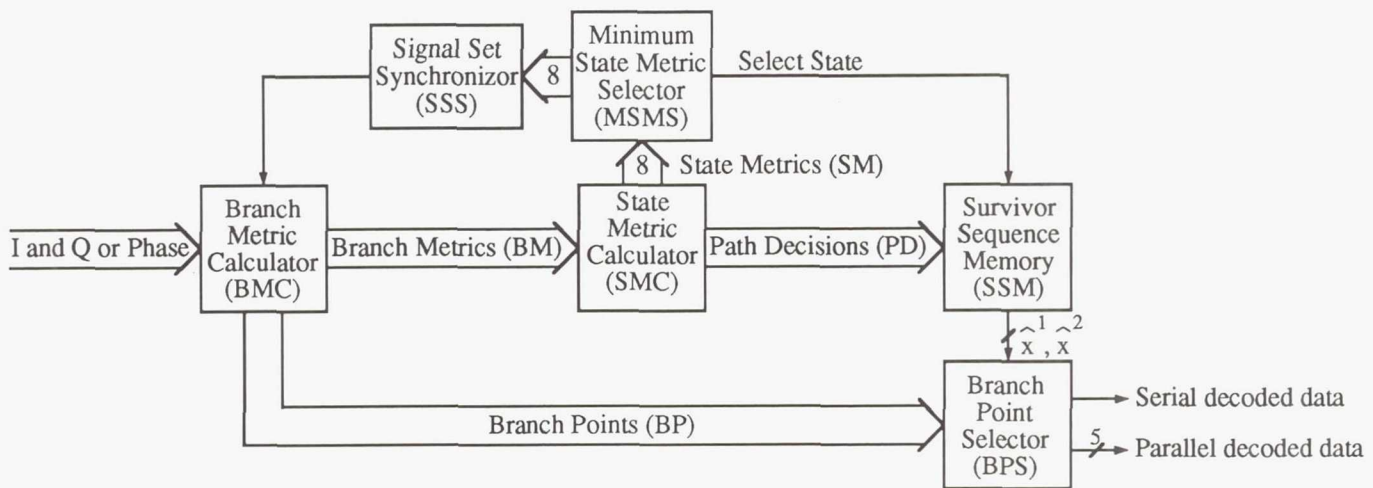
76

Figure 2: Block diagram of a Viterbi decoder for the 16 state 2.5 bit/sym 4D-8PSK trellis code.

the design would work, even though the actual design is twice as complicated. Our design uses a 10 MHz clock (giving 100 ns clock cycles) and Schottky TTL logic for its ease of use and large variety of functions. The actual technologies used are 74LS (Low-power Schottky TTL) for non-time critical sections of the circuit and 74F (Advanced Schottky TTL) for time critical sections. Other technologies are used for functions not available in 74F or 74LS.

The decoder is operated asynchronously to the received data clock. This requires one of the seven extra clock cycles described above. Internally, the decoder operates synchronously to the 10 MHz clock. The decoder starts operation after detecting the first rising edge of the received 4D symbol clock. After 23 clock cycles, the decoder stops and waits for the next rising edge of the 4D symbol clock. This allows the decoder to operate at any data rate from 0 to 2.1 Mbit/s.

Each iteration of the Viterbi algorithm decodes five bits for each received 4D signal point (since the code rate is 5/6). The maximum 4D symbol rate of the decoder is the internal clock speed divided the number of clock cycles required to decode the five bits, i.e., 4.35×10^5 4D symbols per second. Therefore, the maximum bit rate of the decoder is 2.17 Mbit/s. For the HST, this code could achieve a data rate up to 7.5 Mbit/s. For actual use on the HST, it is intended that the decoder would be implemented on a VLSI chip, where the required decoding speed would be achieved.

There are six main sections in the Viterbi decoder. These are

- Branch Metric Calculator (BMC)
- State Metric Calculator (SMC)
- Survivor Sequence Memory (SSM)
- Signal Set Synchronisor (SSS)
- Minimum State Metric Selector (MSMS)
- Branch Point Selector (BPS)

Figure 2 illustrates a block diagram of the decoder. The above sections are described as follows.

3.1 Branch Metric Calculator

For each transition of the trellis there are 8 parallel paths (due to the three unchecked bits in the encoder). The BMC must determine which of the paths is closest to the received 4D signal point (the Branch Point (BP)) as well as the Branch Metric (BM) for this path. The BM can be calculated in a number of ways. The optimum BM's for AWGN channels with quantisation are log-likelihood metrics [4]. Alternatively, one could make an approximation based on the squared Euclidean distance between the received point and the points along the transitions.

In our design we have chosen to use Read Only Memory's (ROM's) to store the

precalculated BP (three bits are used to represent each parallel path) and BM (based on log-likelihood metrics). The encoder can produce one of eight (i.e., $2^{\tilde{k}+1}$) sets of parallel paths (each containing 8 paths). The BP and BM must be determined for each of these eight sets of parallel paths.

We have chosen four bits to represent the BM value. This gives a BM range from 0 (closest to the received 4D point) to 15 (furthest from the 4D point). Decoder simulations in [5] for another multi-D trellis code indicate that this amount of quantisation results in little performance degradation.

To minimise the number of address bits to the ROM, each received 2D signal point has been quantised to seven bits. After extensive simulations in [5] for a 6D-8PSK trellis code, it was found that pie-chart or angular quantisation results in the least performance degradation (0.2 to 0.3 dB for five bit quantisation). The simulations included the "dartboard" quantisation pattern proposed in [1].

Each ROM therefore has an address space of 14 bits (seven bits for each 2D symbol). The ROM's used for the BMC are 32K×8 27C256's. A total of 6 ROM's were used, two for determining the BP's and four for the eight BM's.

Alternative BMC schemes which exploit the finite length trellis structure of the parallel transitions were also considered. That is, a Viterbi like decoder can be used to decode the parallel transitions. However, their large complexity (in a discrete implementation) led us to choose the simpler ROM look-up method. For a VLSI implementation, though, the trellis decoding method would be preferable due to the flexibility that VLSI provides in designing circuits. Thus, the Viterbi decoder (with the BMC) could be implemented on a single chip.

3.2 State Metric Calculator

The SMC updates the State Metrics (SM) for each state of the code in each iteration of the Viterbi algorithm. A SM is an indication of how close the received sequence is to the closest path of all paths leading into a particular state. Since the code has two checked bits, there are four paths leading into each state (since we choose the closest path among the 8 parallel paths in the BMC). For each of the four paths, we must add the BM for that path to its corresponding SM (also known as the old SM) from the previous iteration. The new SM for the four paths leading into a state is the smallest of these summations. This path is selected and all other paths are eliminated. This is called the Add-Compare-Select (ACS) operation.

With four paths into each state a 4:1 ACS circuit is required. With 16 states in our code, the ACS operation needs to be performed 16 times (explaining the need for 16 clock cycles). The ACS circuit also produces two Path Decision (PD) bits which indicate which of the four paths was chosen. This information is passed to the SSM where it is stored.

Since the decoder operates serially, only one ACS circuit is required. The 16 SM's are stored in two 74AS870 dual 16×4 static Random Access Memory (RAM) chips. Eight bits are used to represent each SM. As shown in [5] for a 6D-8PSK trellis code, this is more than enough bits when two's complement arithmetic is used in the ACS circuit to prevent overflow [4]. Before the first new SM can be calculated, four old SM's are read out from the RAM's. This takes four clock cycles. It takes another two clock cycles to perform the ACS operation. To achieve a slightly higher speed, we could have done the ACS operation in one clock cycle. However, this would have required six comparator chips to find the minimum SM. An increase of one clock cycle and the use of three comparator chips was chosen to decrease the complexity of the design.

Another clock cycle is used to write to the other half of the dual 16×4 RAM's. Since all the read and ACS operations are pipelined, an additional 15 clock cycles are required to write the 15 remaining new SM's. In the next iteration of the algorithm we read from where the SM's were written in the

previous iteration and write to where the old SM's had been stored. The process then repeats.

For the ACS circuit, the appropriate BM's must be added to the correct old SM's. Twelve quad 2:1 multiplexer chips and a copy of the convolutional encoder are needed to accomplish this task.

3.3 Survivor Sequence Memory

The SSM has two tasks. It must store the Path Decisions (PD's) generated by the SMC and "traceback" through the previously stored PD's to determine the final decoded bits for x^2 and x^1. This requires alternating write and read (for the traceback) operations on the memory. The traceback depth is the required number of PD sets (each set consists of 16 two bit PD's) that the SSM must trace back through.

The PD's must be stored in the remaining 16 clock cycles that are available. There are two ways this can be achieved. Storing two PD bits in each clock cycle or storing four PD bits in every other cycle, leaving the alternate cycle to perform part of the traceback. With the first method at least two separate memories are required since the traceback operation cannot be performed simultaneously with the storage of the new set of PD's (due to the design of memory chips). Since there is a finite amount of memory, the oldest PD set must be written over.

There is usually a point where one method is better than the other (in terms of the total memory size required) based on the number of clock cycles available and the traceback depth. A traceback depth of around 25 to 30 results in little performance degradation [5]. Comparing the implementation complexity of the two methods, the alternating read/write method proved superior.

With this design only eight clock cycles are available to perform a traceback. To maintain integer power of 2 address spaces for the memories (and thus efficiently use of practical memory designs), a traceback depth of seven is used for each SSM memory chip. To achieve the required traceback depth, four 64×4 memories are required. This gives a traceback depth of 28. The traceback is performed in a pipelined fashion, switching between memories when required and waiting for the next received set of data to continue with the traceback. Four separate memories are required since there are four tracebacks in operation at any one time.

Since there are no 64×4 RAM's commercially available, larger 256×4 93422A RAM's were used. This chip has separate input and output data buses which simplifies the SSM design. We use the state with the smallest SM to start the traceback. This is the best state the SSM could start with (since it corresponds to the path that is closest to the received signal) and helps give the decoder a slight performance improvement over choosing a random or a fixed state. The Minimum State Metric Selector (MSMS) provides the information needed to achieve this.

At the correct time and place in the circuit, the two decoded bits x^1 and x^2 are produced. The two bits are passed to the Branch Point Selector (BPS) where they are re-encoded to select one of the eight 3 bit branch points. The branch points are delayed by 34 4D symbol periods, 28 due to the traceback, 4 due to the pipeline delay in the traceback, and 2 due to the re-encoding of the decoded data.

The five decoded bits are then differentially decoded (optional) and then parallel to serial converted for the final decoder output. Precoding and postdecoding are optional as there are some communication systems that do not require phase synchronisation. For example, a burst modem can provide phase information in the preamble of a burst. A 74HC4046 PLL is used to generate the required five times clock for the serial data. This PLL is tuned to lock within 0 to 2 MHz, but as expected for PLL's the lower frequency limit will be somewhat greater than DC. The decoded data is also available in five bit bytes.

3.4 Signal Set Synchroniser

The SSS has the task of synchronising the decoder to the received sequence of 2D symbols. Since the signal set consists of two 2D signals, the decoder must synchronise to one of the two possible ways the received data can arrive.

The decoder is asynchronously locked to DATCLK, which is the received 2D symbol clock whose frequency has been divided by two. A delay of zero or one 2D symbol periods of DATCLK is used for timing synchronisation.

The SSS works by examining the rate of increase of the minimum SM from the MSMS. If the rate is high, this indicates that the decoder is out of synch and needs to be resynchronised. A variable threshold in the SSS is used for this purpose. If the threshold is exceeded, the SSS will toggle into the "arm symbol toggel" state.

If the threshold is again exceeded in the next V (V is a variable from 0 to 63) 4D symbol periods the decoder will toggle the 2D symbol delay (from zero to one or one to zero). The SSS then ignores the decoder for 128+V 6D symbol periods to allow the decoder to settle into its new signal set configuration.

If the threshold is not exceeded the SSS will "disarm" and return to its normal monitoring state.

4 OTHER DECODER FEATURES

The encoder and decoder are mounted within a 3U high 19 inch rack. On the front panel, two Light Emitting Diodes (LED's) are used to indicate the 2D symbol delay.

To test the decoder, the 2D symbol delay can be independently set to manual control. In this way, the SSS can be isolated from the rest of the circuitry so that any problems with the rest of the decoder can be fixed without the SSS interfering. It can also be used to test the SSS by manually introducing delays into the received signal. There are two switches used for this.

Two rotary type switches are used to select the format of the received data. One switch is used to select between 3 bit phase (corresponding to hard decision), 7 bit phase quantisation, 5 bit I and Q quantisation, or internal loopback mode. The other switch selects between signed magnitude, reverse binary, straight binary, or two's complement data formats for I and Q received data.

There are also switches for disabling the postdecoder from the decoder and the precoder from the encoder. The encoder has another switch to select between five bit parallel or bit serial data. The decoder also has a reset button to force all the SM's to zero. The encoder/decoder interface diagram is given in Figure 3.
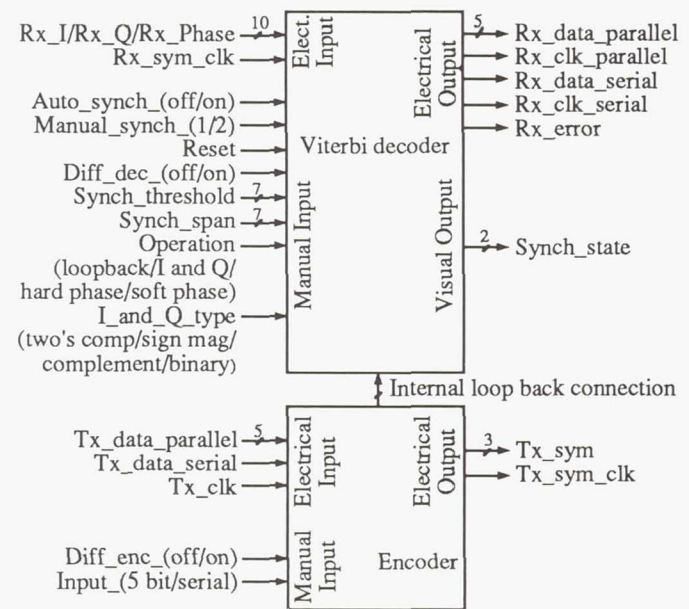
Figure 3: Viterbi decoder/encoder interface diagram for 16 state 2.5 bit/sym 4D-8PSK trellis code.

The 159 integrated circuits of the design are placed on two double height Speedwire Eurocards (233.4×220 mm). Speedwire allows quick and reliable connections (if it is done correctly) between the chips that can be easily changed. The speedwire boards also have good groundplanes, critical when operating at high clock speeds. The Viterbi decoder (which operates at 10 MHz) is placed on one board (taking 96 chips) while the encoder, SSS, and various interface chips are placed on the

other board.

BNC connectors are used at the back of the rack for external data and clock connections. It is assumed that all received data changes on the rising edge of its clock. Similarly, the codec produces its signals in the same format. TTL 75 Ω interface signals are used for these external interfaces.

6 CONCLUSIONS

A serial implementation of a Viterbi decoder for the 16 state 2.5 bit/sym code with a 4D-8PSK signal set has been described. This decoder can provide high data rates (up to 2.1 Mbit/s) and is intended for future use on the Hubble Space Telescope. Due to its serial implementation the decoder design is quite complex, but could be implemented on a single VLSI integrated circuit.

The Branch Metric Calculator has been implemented through the use of large look-up table ROM's. A VLSI implementation may use a Viterbi type decoding algorithm to allow single chip implementation.

REFERENCES

[1] Pietrobon, S. S., R. H. Deng, A. Lafanechère, G. Ungerboeck, and D. J. Costello, Jr., "Trellis-coded multidimensional phase modulation," *IEEE Trans. Inform. Theory*, vol. 36, pp. 63-89, Jan. 1990.

[2] Perez, L., "On the performance of multi-dimensional phase modulated trellis codes," NASA Tech. Report #89-10-02, Oct. 1989.

[3] Porath, J. E., "Algorithms for converting convolutional codes from feedback to feedforward form and vice versa," *IEE Electron. Lett.*, vol.25, pp. 1008-1009, 20 Jul. 1989.

[4] Pietrobon, S. S., "Rotationally invariant convolutional codes for MPSK modulation and implementation of Viterbi decoders," M.Eng. Thesis, School of Electron. Eng., South Australian Inst. of Technol. (now University of South Australia), Adelaide, Australia, Jun. 1988.

[5] Gray, P. K., I. S. Morrison, and W. G. Cowley, "The development of a 45 Mbit/s modem/codec," *Proc. IREECON-'89*, pp. 942-945, Melbourne, Australia, Sep. 1989.

Page intentionally left blank

AN OVERVIEW OF SPACE COMMUNICATION ARTIFICIAL INTELLIGENCE FOR LINK EVALUATION TERMINAL (SCAILET) PROJECT

Anoosh K. Shahidi
Sverdrup Technology, Inc.
Lewis Research Center Group
Brook Park, Ohio

Richard F. Schlegelmilch
The University of Akron
Akron, Ohio

Edward J. Petrik and Jerry L. Walters
NASA Lewis Research Center
Cleveland, Ohio

ABSTRACT

A software application to assist end-users of the high burst rate (HBR) link evaluation terminal (LET) for satellite communications is being developed. The HBR LET system developed at NASA Lewis Research Center is an element of the Advanced Communications Technology Satellite (ACTS) Project.

The HBR LET is divided into seven major subsystems, each with its own expert. Programming scripts, test procedures defined by design engineers, set up the HBR LET system. These programming scripts are cryptic, hard to maintain and require a steep learning curve. These scripts were developed by the system engineers who will not be available for the end-users of the system.

To increase end-user productivity a friendly interface needs to be added to the system. One possible solution is to provided the user with adequate documentation to perform the needed tasks. With the complexity of this system the vast amount of documentation needed would be overwhelming and the information would be hard to retrieve. With limited resources, maintenance is another reason for not using this form of documentation.

An advanced form of interaction is being explored using current computer techniques. This application, which incorporates a combination of multimedia and artificial intelligence (AI) techniques to provided end-users with an intelligent interface to the HBR LET system, is comprised of an intelligent assistant, intelligent tutoring, and hypermedia documentation. The intelligent assistant and tutoring systems address the critical programming needs of the end-user.

INTRODUCTION

A software application to assist end-users of the link evaluation terminal (LET) for satellite communications is being developed. This software application incorporates artificial intelligence (AI) techniques and will be deployed as an interface to LET. The high burst rate (HBR) LET provides 30 GHz transmitting/20 GHz receiving, 220/110 Mbps capability for wideband communications technology experiments with the Advanced Communications Technology Satellite (ACTS). The HBR LET can monitor and evaluate the integrity of the HBR

communications uplink and downlink to the ACTS satellite. The uplink HBR transmission is performed by bursting the bit-pattern as a modulated signal to the satellite. The HBR LET can determine the bit error rate (BER) under various atmospheric conditions by comparing the transmitted bit pattern with the received bit pattern. An algorithm for power augmentation will be applied to enhance the system's BER performance at reduced signal strength caused by adverse conditions.

The HBR LET terminal consists of seven major subsystems:

- Antenna subsystem
- Radio frequency (RF) transmitter subsystem
- RF receiver subsystem
- Control and performance monitor (C&PM) computer subsystem
- Local loopback subsystem at RF
- Modulation and BER measurements subsystem
- Calibration subsystem

The C&PM computer controls and monitors all the other subsystems through an IEEE488 interface. HBR LET experiments with the ACTS satellite will be initiated by users through the C&PM experiment control and monitor (ECM) software. The ECM software was developed on a Concurrent 3205 minicomputer in FORTRAN, which provides the end-user with the following capabilities:

- Individual instrument control
- Interactive interface used to communicate with the digital ground terminal
- Ability to conduct BER measurements
- User-controlled data acquisition

Programming scripts, defined by the design engineer, set up the HBR LET terminal by programming subsystem devices through IEEE488 interfaces. However, the scripts are difficult to use, require a steep learning curve, are cryptic, and are hard to maintain. The combination of the learning curve and the complexities involved with editing the script files may discourage end-users from utilizing the full capabilities of the HBR LET system.

The following SCAILET features will improve the HBR LET system and enhance the end-user's ability to perform the experiments:

INTELLIGENT ASSISTANT

An intelligent assistant is a software program that will aid the user in operating of the HBR LET components. Friendly human interfaces shield the user from the script and complexities of the HBR LET system and furthermore aid performing iterative setup tasks. Any intelligent assistant also contains sufficient information about the HBR LET system to alert the user to erroneous actions.

The intelligent assistant uses a personal computer to provide a dynamic schematic diagram of the overall system . The schematic diagram provides a graphic user interface which serves as a "front-end" to the HBR LET system. The user can control the system by interacting with the schematic diagram. An expert system handles requested changes in the system. The changes will then be reflected dynamically on the schematic diagram.

The dynamic graphics are implemented using the Choreographer Graphical User Interface toolkit, developed by GUIdance Technologies Corp. The expert system shell used for this project is KAPPA PC developed by Intellicorp Corp. KAPPA PC is a hybrid expert system shell which has both a complete object oriented programming system and a rule based expert system. All HBR LET devices are implemented in KAPPA PC objects. The devices are connected to one another using message passing to create a dynamic model of the system. Rules are used to create programming scripts.

MULTIMEDIA DOCUMENTATION

Multimedia is a software package that applies a combination of text, graphics, voice, and video technologies in a computer application tool. The combination of these technologies provides a more powerful tool than any one medium alone. Text and graphics information can then be combined within the application. Similarly, associated voice and video information can be integrated into the application. This module is implemented using the ToolBook hypertext shell developed by Asymetrix Corp.

The documentation for the HBR LET subsystems is being developed by different design engineers. However, the end-user will need to see the association among the subsystems. Multimedia allows the user to look at a graphic image of a schematic diagram, and request written specifications of the components, more detailed images, or actual assembly diagrams. Multimedia allows the users to follow schematics' links and visit subsystems within the circuitry. Using printed documents would be less friendly and require much more time and effort to understand. This module is connected to the intelligent assistant so the user can access relevant documentation anytime during system programming.

FUTURE DIRECTION: INTELLIGENT TUTOR

Computer assisted instruction (CAI) is a traditional computer based training program which takes the user through a predetermined set of lessons. The advent of artificial intelligence technology and advances in cognitive psychology gave rise to intelligent tutoring systems (ITS) as an improvement to CAI. In an ITS environment, the curriculum designer determines what concepts the student should learn in a lesson. The student is then taken through subjects to see which concepts he/she is lacking. The program then determines the curriculum depending on the needs of the student.

For HBR LET, an initial overview of individual system components is necessary to aid in understanding the complete system. Concepts important to the operation of each HBR LET subsystem will be identified for SCAILET after the multimedia documentation and intelligent assistant tasks are completed . A guided learning process, which incorporates the use of a simulator, will be developed to provide ITS instruction on the operation of the HBR LET system.

ACKNOWLEDGMENTS

*Principle contact: Anoosh K. Shahidi, Sverdrup Technology Inc. Mail Stop SVR-1, 2001 Aerospace Parkway, Brook Park, Ohio 44142 (216) 891-2213

BIBLIOGRAPHY

1. "Using Open Script: ToolBook Software Construction Set for Windows". Asymetrix Corp., Bellevue, Washington 98004. (1990)

2. "KAPPA PC Reference Manual". Intellicorp Corp., Mountain View, CA (1989).

3. " Choreographer Reference Manual". Guidance Technologies Corp. (1989)

A RECONFIGURABLE MULTICARRIER DEMODULATOR ARCHITECTURE

S. C. Kwatra and M. M. Jamali
Department of Electrical Engineering
University of Toledo
Toledo, Ohio 43606
419-537-2060

ABSTRACT

An architecture based on parallel and pipeline design approaches has been developed for the FDMA/TDM conversion system. The architecture has two main modules namely the transmultiplexer and the demodulator. The transmultiplexer has two pipelined modules. These are the shared multiplexed polyphase filter and the FFT. The demodulator consists of carrier, clock and data recovery modules which are interactive. Progress on the design of the MCD using commercially available chips and ASIC and simulation studies using Viewlogic software will be presented at the conference.

INTRODUCTION

Presently most satellite communication systems require a large earth station to supply a network with multiplexed speech and video channels. In the future, satellite communication systems will consist of a large number of small capacity, multi-service users. For these systems the conventional transmission methods of FDMA or TDMA access are no longer efficient. One approach to offer these services at a low cost to the user is to use SCPC/FDMA on the uplink and TDM on the downlink [1-3]. The problem with this type of communication is that it transfers the burden of computation on-board the satellite, where power and area requirements are critical. It can thus be seen that hardware that is efficient in terms of speed, power consumption and components needs to be developed for performing the computations on board the satellite. To perform the FDMA/TDM conversion , a Multicarrier Demodulator (MCD), baseband switch matrix, TDM multiplexer and modulator are required on board the satellite. The MCD consists of a transmultiplexer followed by a bank of demodulators. The transmultiplexer is required to separate the FDMA signal into individual channels. The bank of demodulators take the separated channels and recover the data from them.

To take advantage of the developments in the area of VLSI and digital systems, a digital implementation of the reconfigurable MCD has been proposed [3,5]. The algorithm selected for performing the transmultiplexing is the polyphase FFT method. The polyphase FFT algorithm consists of a filter bank followed by a FFT operation [4,9]. A PROgrammable DEModulator (PRODEM) which uses a single shared device to demodulate all the channels has also been proposed [5]. The demodulator consists of three modules namely carrier recovery, clock recovery and data recovery [6-8].

SYSTEM DESIGN

Quadrature sampling is used to digitize the analog signal so that the components used in the architecture can operate at a lower rate and also because the complex representation is compatible with the FFT. The RTMUX is capable of demultiplexing channels in three different cases. The three cases are :
1) 800 channels at 64 Kb/s each.
2 (a) A mix of 400 channels at 64 Kb/s and 2 (b) 12 channels at 2.048 Mb/s each.
3) 24 channels at 2.048 Mb/s each.
Since case 2 is a mix of two carriers, each carrier having its own demultiplexing characteristics, it is split into its two constituent parts 2(a) and 2(b) so that each case can then be demultiplexed individually.

The system diagram of the Reconfigurable Transmultiplexer (RTMUX) that is capable of demultiplexing the above three cases is shown in Fig. 1. The front end of the RTMUX consists of a demultiplexer that routes the input FDMA signal to one of the three modules, namely, modules 1, 2 and 3 whose functions are described below. Since cases 2(a) and 2(b) are part of case 2 and each occupies half of the total spectrum, they need to be split into two halves. The individual channels can then be demultiplexed from the two halves. Module 1 is used to split the input FDMA signal into two halves for cases 2(a) and 2(b). Module 2 is designed to demultiplex either case 1 or case 2(a) by reconfiguring itself. Similarly module 3 is designed to demultiplex either case 2(b) or 3 by reconfiguring itself. A Reconfigurable Shared Filter Bank has been designed and is shown in Fig. 2. The FFT pipeline proposed in [3,4] is made reconfigurable by varying the number of FFT stages in the pipeline and by having a programmable FFT coefficient and address generator. A N-1/N stage RFFT that can perform either a 2^{N-1} or 2^N FFT is shown in Fig. 3. A multiplexed AE (MAE), to implement the FFT butterfly operation is designed along with its interface to the other components in the RFFT pipeline and is shown in Fig. 4.

Programmable Demodulator (PRODEM)

The PRODEM consists of three main modules namely Multiplexed Carrier Recovery Module (MCRM), Multiplexed Timing Recovery Module (MTRM) and Multiplexed Data Recovery Module (MDRM). The hardware is constructed to demodulate all the channels simultaneously. The number of channels can be varied as long as the total bit rate is below the maximum bit rate sustained by the modules. Moreover the bit rate of a group of channels could also vary.

A MCRM is designed to obtain the carrier phase for each of the channels. Samples of several channels are input serially to this module. At the same time these samples are also input to the Multiplexed RAM Buffer for Samples (MRBS) which stores these samples to be operated on later by the phase recovered information of the MCRM. The output of the MCRM module will be needed by the MDRM. The in-phase and quadrature-phase samples of the channels are input to the MCRM as shown in Fig. 5.

A MRBS is designed to store the incoming samples for the duration of an estimation interval and is shown in Fig. 6. The MCRM operates on the samples obtaining the carrier phase for the channels. Also, at this time the input samples are buffered in the MRBS. The MDRM uses the output of the MCRM along with the stored values of MRBS to recover the digital data. This design uses a single RAM-Latch combination at each stage to store samples of different channels corresponding to each AGC cycle.

A MDRM is designed to extract the digital information. It operates on the samples processed by the MCRM and MRBS. The hardware design is shown in Fig. 7. The MDRM module utilizes the in-phase and quadrature-phase samples from the MRBS along with the sine and cosine values of the MCRM to extract the digital data for all the channels. At any time four values are input to this module. The output is computed and stored in a latch preceeding the Digital Data RAM (DDR). Also, these values are used as an input to the MTRM. After the necessary computations are performed the results are stored in unique locations of the Digital Data RAM (DDR).

A MTRM is designed to extract the timing information needed for tracking the input samples and is shown in Fig. 8. This timing information is used by the interpolator. Its input is available from the latches used preceeding the DDR of the MDRM. The output of these latches is used as an input to the MTRM. The combination of the four modules namely MCRM, MRBS, MDRM and MTRM is collectively called a PRODEM. These four modules need to be appropriately interfaced. The addressing scheme, control circuitry and the integration of addressing units for all the modules need special

attention. A design for each of the modules with proper interfaces is shown in Fig. 9.

CONCLUSION AND FUTURE DIRECTIONS

In this paper parallel, pipeline and time sharing concepts are used in the design of a digital MCD. The MCD consists of the TMUX and the PRODEM. The TMUX is implemented by means of a shared filter bank module and a pipelined FFT module. The PRODEM consists of three modules namely carrier, clock and data recovery modules, which have been optimized for high speed demodulation. The shared filter bank and PRODEM process the channels in a time multiplexed manner. At present we are designing both TMUX and PRODEM both at the board and ASIC levels.

REFERENCES

[1]   G.Perrota and G.Losquadro, "FDMA/TDM satellite communication systems for domestic/business services", Proceedings International Conference on Digital Satellite Communication -7, pp. 155-162, 1986.
[2]   B.G.Evans and F.P.Coakley, "Multi carrier demodulators for OBP satellites", International Journal of Satellite Communication, Vol 6, pp. 243-251, 1988.
[3]   J.Campanella and S.Sayegh, "Onboard multichannel demultiplexer/demodulator", NASA contract no cr 1801827 , 1987.
[4]   P.J.Fernandes, "A Reconfigurable Transmultiplexer Architecture", M.S.Thesis, University of Toledo, 1991.
[5]   L.P.Eugene, "A Parallel Pipelined Architecture for a Multi Carrier Demodulator", M.S.Thesis, University of Toledo, 1991.
[6]   E. Del Re and R. Fantacci, "Alternatives for onboard digital multicarrier demodulation," International Journal of Satellite Communications, Vol. 6, pp. 267-281, 1988.
[7]   A. J. Viterbi and A. M. Viterbi, "Non-linear estimator of PSK modulated carrierphase with application to burst digital transmission," IEEE Transactions on Information Theory, IT-29, pp. 543- 551, 1983.
[8]   F. M. Gardner, "A BPSK /QPSK timing error detector for sampled receivers," IEEE Transactions on   Communications, COM-34, pp. 423-429, 1986.
[13]   M. Bellanger and P. Daguet , "TDM/FDM transmultiplexers: digital polyphase and FFT," IEEE Transactions on Communication , pp. 1199-1205, Sep. 1974.
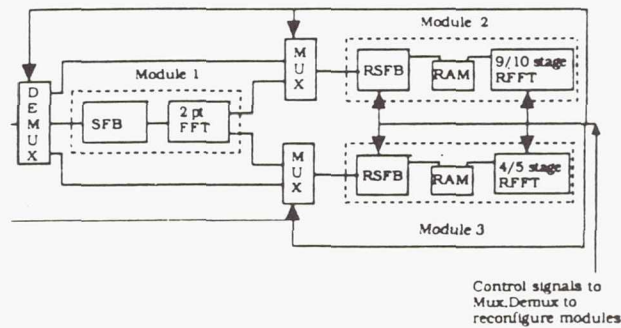
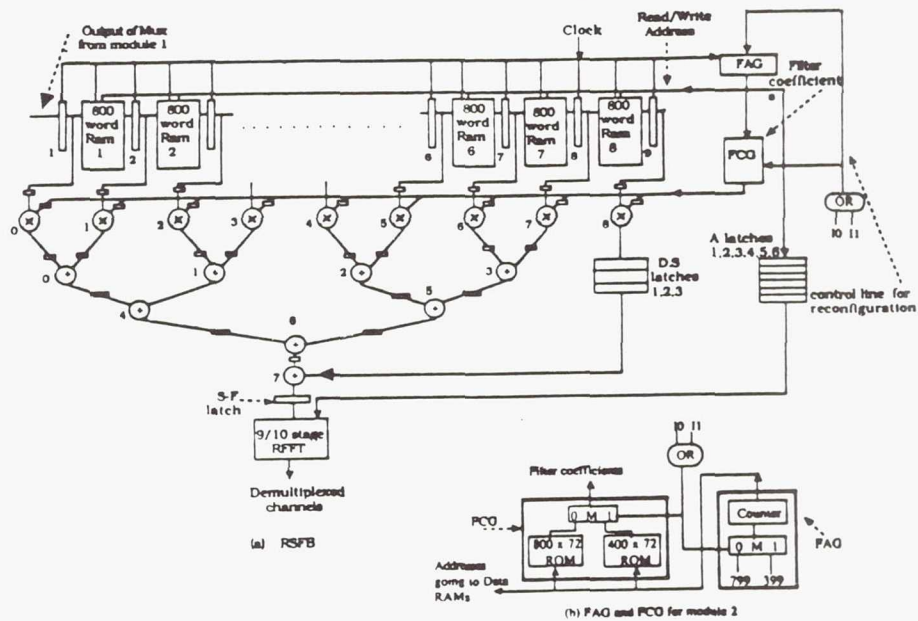Fig. 1. Reconfigurable Transmultiplexer (RTMUX)

(a) RSFB

(b) FAG and FCG for modules 2

Fig. 2. Reconfigurable Shared Filter Bank (RSFB)

(a) RFFT pipeline
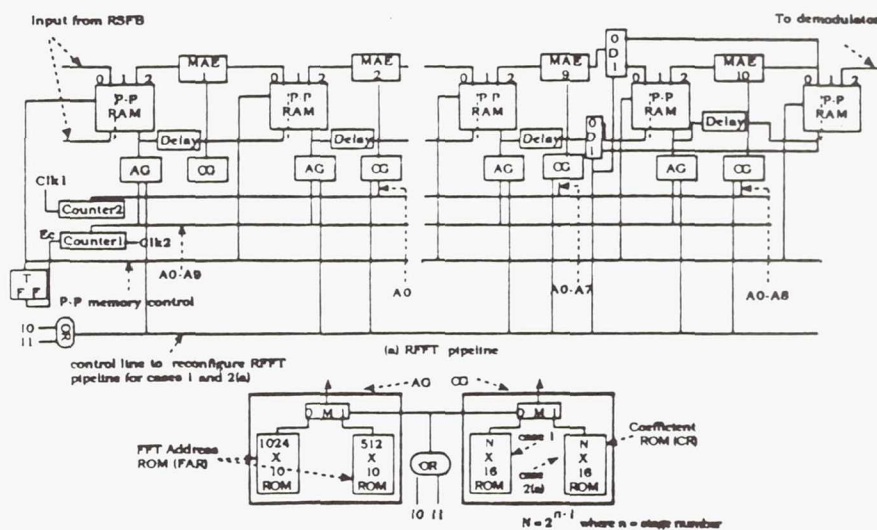
Fig. 3. N-1/N Stage RFFT
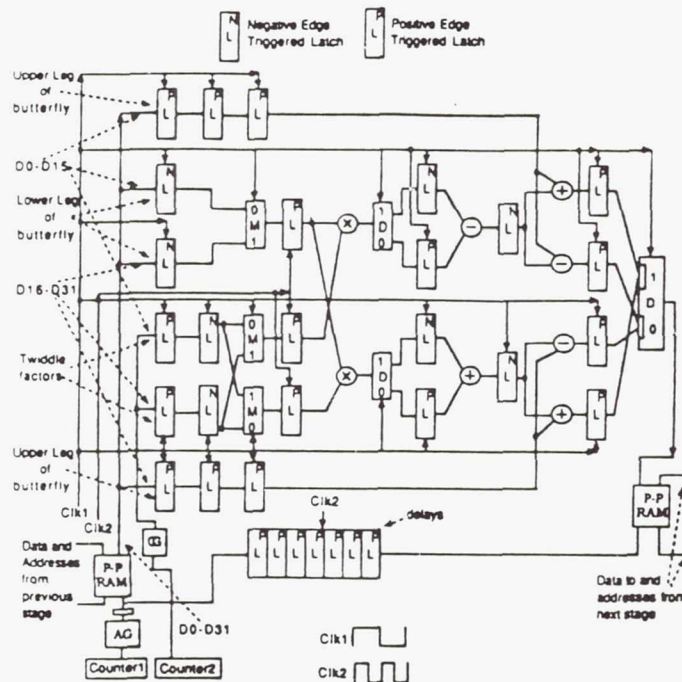
Fig. 4. Detailed Implementation Structure of the MAE in the RFFT

AGC - ADDRESS GENERATOR CHANNEL
AGS - ADDRESS GENERATOR SAMPLE
AR - ACCUMULATION RAM
IR - INPUT ROM
RE - READ ENABLE
OR - OUTPUT ROM
SR - STORAGE RAM
Tccd - TERMINAL COUNT FROM AGC
Tcs - TERMINAL COUNT FROM AGS
WE - WRITE ENABLE

Fig. 5. Multiplexed Carrier Recovery Module (MCRM)

AGS - ADDRESS GENERATOR SAMPLES
L - LATCH
O - READ ENABLE
R - RAM
W - WRITE ENABLE
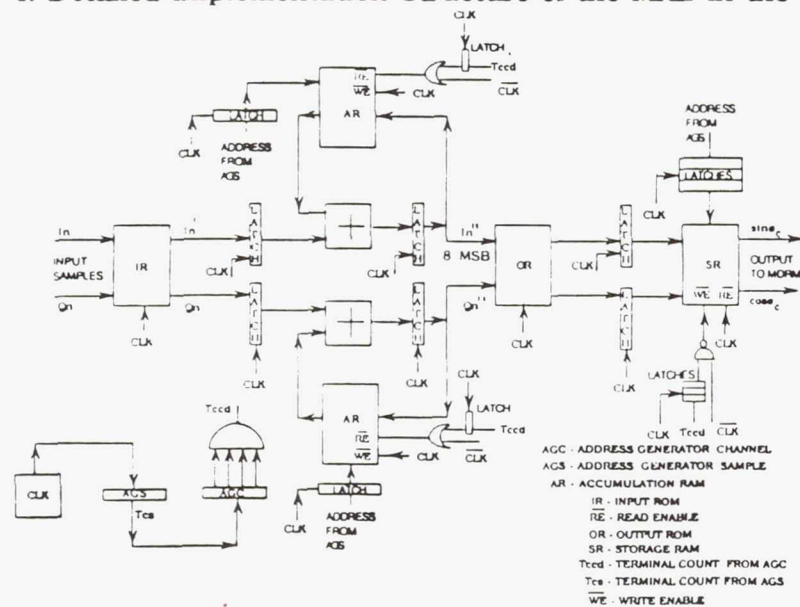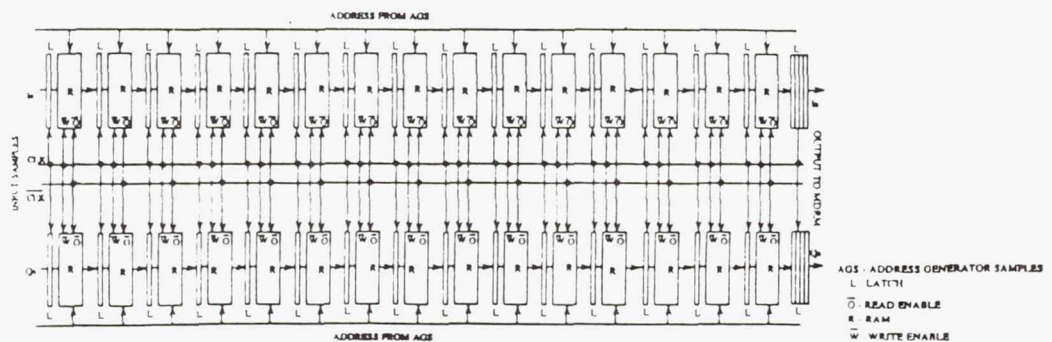
Fig. 6. Multiplexed RAM Buffer for Samples (MRBS)

92

Fig. 7. Multiplexed Data Recovery Module (MDRM)

AGC - ADDRESS GENERATOR CHANNEL
AGS - ADDRESS GENERATOR SAMPLE
DDR - DIGITAL DATA RAM
OE - READ ENABLE
Tcc - TERMINAL COUNT FROM AGC
Tcs - TERMINAL COUNT FROM AGS
WE - WRITE ENABLE

Fig. 8. Multiplexed Timing Recovery Module (MTRM)

SR - STORAGE RAM
AR - ACCUMULATION RAM
DDR - DIGITAL DATA RAM
TR - TIMING RAM
IR - INPUT ROM
OR - OUTPUT ROM

Fig. 9. PRODEM

93
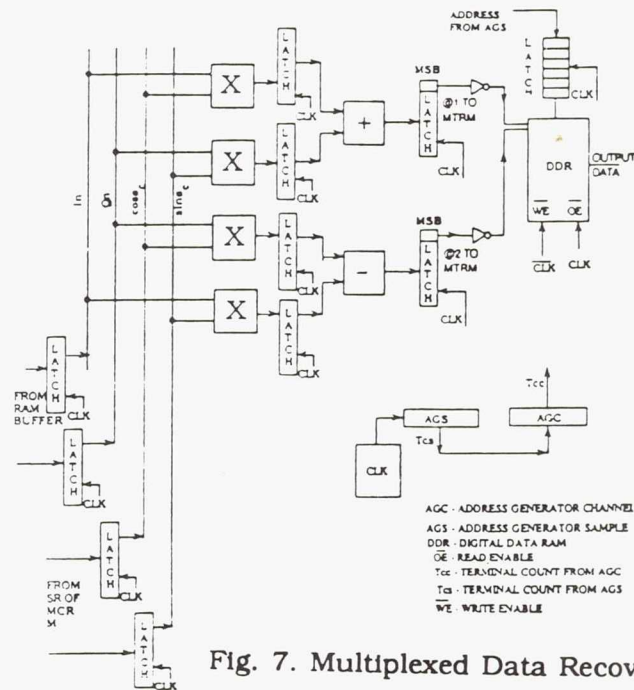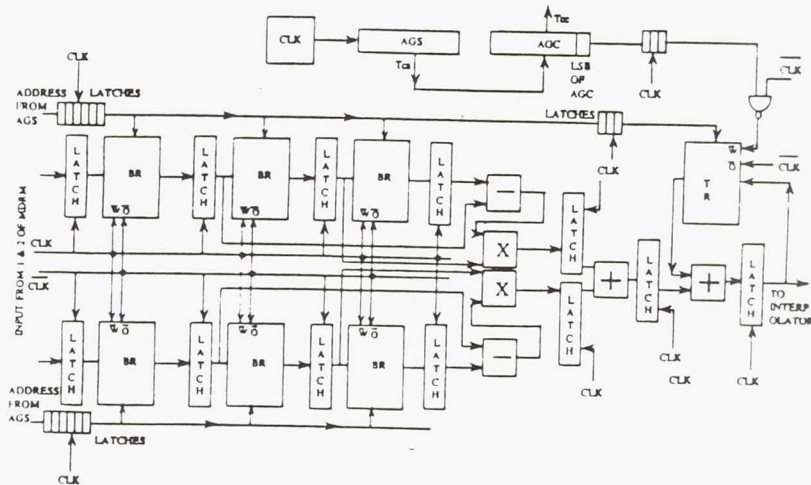
Page intentionally left blank

Video Data Compression Using Artificial Neural Network Differential Vector Quantization

Ashok K. Krishnamurthy   Steven B. Bibyk   Stanley C. Ahalt

Department of Electrical Engineering
The Ohio State University
Columbus, Ohio 43210

Abstract

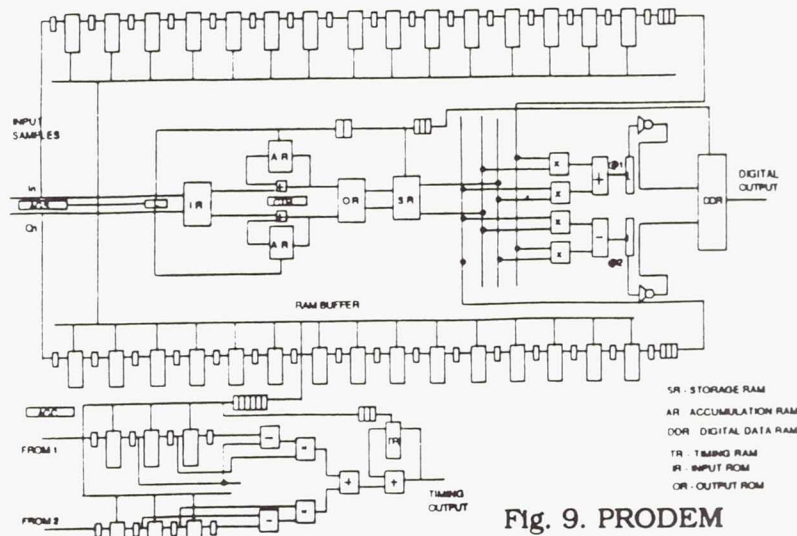An artificial neural network vector quantizer is developed for use in data compression applications such as Digital Video. Differential Vector Quantization is used to preserve edge features, and a new adaptive algorithm, known as Frequency-Sensitive Competitive Learning, is used to develop the vector quantizer codebook. To develop real time performance, a custom VLSI ASIC is being developed to realize the associative memory functions needed in the vector quantization algorithm. By using vector quantization, the need for Huffman coding can be eliminated, resulting in superior performance against channel bit errors than methods that use variable length codes.

1   Introduction

Effective data compression algorithms are needed to reduce transmission bandwidth and storage space. In particular, there is a great deal of interest in the low bit rate coding of images. In this paper, we discuss the compression of digital video image data, which has become a central concern as HDTV standards begin to develop. One compression technique, Vector Quantization (VQ) [1, 2], has emerged as a powerful technique that can provide large reductions in bit rate while preserving essential signal characteristics. In this paper we show that error-insensitive VQ encoders can be constructed by employing *entropy* based VQ codebooks.

The purpose of this paper is to describe the use and implementation of an Artificial Neural Network (ANN) Vector Quantizer. More specifically, we discuss the design of a real-time, edge-preserving Differential Vector Quantizer (DVQ) architecture. We discuss the use of an ANN algorithm to design VQ codebooks, and we anticipate that the use of the same ANN algorithm can be employed in *adaptive* DVQ coders. The particular ANN algorithm we use is called Frequency Sensitive Competitive Learning (FSCL). This algorithm

has been described in depth in previous publications [3, 4], so only a brief presentation is given here.

A locally-optimal vector quantization algorithm, proposed by Linde, Buzo, and Gray (LBG) [5], has been extensively employed in encoding both speech and images. However, studies have shown that, in many cases, the computational complexity of this algorithm restricts its use in real-time applications [1, 6]. The use of ANNs to perform vector quantization has been proposed to overcome these limitations.

The use of ANNs for vector quantization has a number of significant advantages. First, ANNs are highly-parallel architectures and thus offer the potential for real-time VQ. Second, the large body of training techniques for ANNs can be adapted to yield new, and possibly better, algorithms for VQ codebook design. Third, in contrast to the batch training mode of algorithms based on the LBG algorithm [7], most ANN training algorithms are adaptive; thus, ANN based VQ design algorithms can be used to build *adaptive* vector quantizers [8]. This is crucial in applications where the source statistics are changing over time.

This paper is organized as follows. First, we briefly describe basic Vector Quantization techniques and discuss ANN VQ techniques. We then describe the FSCL algorithm and show how the FSCL algorithm attempts to build a maximum-entropy codebook. Then, in Section 3, we describe how a VQ encoder can be viewed as an Associative Memory (AM) and discuss issues related to the design and implementation of an AM. This is followed by a short discussion on the Differential Vector Quantization architecture which is used to minimize edge distortion. We then present our experimental results in Section 5 where a FSCL codebook is used in a DVQ architecture to compress digital images.

2 Vector Quantization and the FSCL Artificial Neural Network

2.1 Basic Vector Quantization Concepts

Vector quantization capitalizes on the underlying structure of the data being quantized. The space of the vectors to be quantized is divided into a number of regions of arbitrary volume and a reproduction vector is calculated for each region. Given any data vector to be quantized, the region in which it lies is determined and the data vector is then represented by the reproduction vector for that region. Instead of transmitting or storing a given data vector, a symbol which indicates the appropriate reproduction vector is used. This can result in considerable savings in transmission bandwidth, albeit at the expense of some distortion.

More formally, vector quantization maps arbitrary data vectors to a binary representation or symbol. Thus, the VQ mapping is from a k–dimensional vector space to a finite set of symbols, $\mathbf{M}$. Associated with each symbol $m \in \{\mathbf{M}\}$ is a reproduction vector $\hat{\mathbf{x}}_m$. The encoding of the data vector $\mathbf{x}$ to the symbol m is a mapping,

$$VQ : \mathbf{x} = (x_1, x_2, \cdots, x_k) \rightarrow m$$

where $m \in \{\mathbf{M}\}$ and the set $\mathbf{M}$ has size M. Assuming a noiseless transmission or storage channel, m is decoded as $\hat{\mathbf{x}}_m$, the reproduction vector associated with the symbol m. The collection of all possible reproduction vectors is called the *reproduction alphabet* or more commonly the *codebook*. Since there are M elements in the set $\mathbf{M}$, there are M possible entries in the codebook. Once the codebook is constructed and, if necessary, transmitted to the receiver, the encoded symbol m acts as an index into the codebook. Thus, the *rate*, R, of the quantizer is $R = \log_2 M$ bits per input vector. Since each input vector has k components, the number of bits required to encode each input vector component is R/k.

Since each data vector must be ultimately represented as one of the codebook entries, the composition of the codebook determines the overall performance of the system. A number of different performance criteria can be used to determine an optimal codebook. For example, in image transmission applications the usual objective is to minimize the overall distortion in the signal due to VQ. Thus the design criterion used to design an optimal codebook is the minimization of the average distortion in encoding vectors using the codebook. Another possible criterion is to maximize the entropy of the codebook, i.e., to ensure that each of the codewords is used equally frequently in encoding

the data. This is a very useful criterion in developing ANN training algorithms for VQ design because maximum entropy codebooks can be employed without the use of Huffman codes, thus reducing encoder sensitivity to channel errors. Finally, an alternative criterion is to use a distortion measure that incorporates expected responses of the human visual system to differences in intensity values and motion.

Given a performance criterion, the VQ codebook design process involves the determination of a codebook that is optimal with respect to this criterion. This normally requires knowledge of the probability distribution of the input data. Typically, however, this distribution is not known, and the codebook is constructed through a process called *training*. During training, a set of data vectors that is representative of the data that will be encountered in practice is used to determine an optimal codebook.

During the training process, a distortion measure, $d(\mathbf{x}, \hat{\mathbf{x}})$ is typically used to determine which data points are to be considered as being in the same region. The distortion measure can be viewed as the cost of representing $\mathbf{x}$ as $\hat{\mathbf{x}}$. By determining which training data vectors lie in the same region, the k–dimensional data space is partitioned into cells. All of the input vectors that fall into a particular cell are mapped to a single, common reproduction vector.

2.2 Motivations for the use of ANN VQs

Unfortunately, the VQ training and encoding processes are computationally expensive. Moreover, most of the algorithms currently used for VQ design are *batch mode* algorithms [5], and need access the entire training data set during the training process. Using ANN adaptive techniques, it is possible to realize an adaptive VQ coder in which codewords are modified based on the arrival of each new training vector.

2.3 The FSCL Algorithm

The Frequency - Sensitive Competitive Learning (FSCL) algorithm is an unsupervised ANN consisting of two layers. The input layer nodes transmit the input vector elements to each of the nodes in the output layer. In the output layer, known as the *winner-take-all layer*, each node receives inputs from all of the input nodes. The weighted interconnections between these two layers are considered the *exemplar*, or *weight* vectors and are used for selecting the *winner node*. The winning node is selected on the basis of a modified distortion measure for each of the output layer nodes. The FSCL codebook design algorithm used in a training phase is discussed below.

96

One of the motivations for the Frequency-Sensitive Competitive Learning (FSCL) network is to overcome the limitations of simple competitive learning network while retaining its computational advantages. One of the main problems with CL networks is that some of the neural units may be under-utilized, the learning algorithm for the FSCL network keeps a count of how *frequently* each neural unit is the winner. This information is used to ensure that, during the training process, all neural units are modified an approximately equal number of times. This yields a codebook that, on average, utilizes all of codewords equally. Consequently, the use of variable-length Huffman codes is unnecessary because *no additional compression will be* achieved through their use.

To solve the under-utilization problem and obtain an equiprobable codebook, the FSCL Algorithm uses a fairness function, $\mathcal{F}(u_i)$, which is a function of the local *update counter*, u_i, and is chosen to ensure the utilization of all the nodes in the winner-take-all layer. The motivation and use of fairness function has been discussed in previous papers [3, 4].

Finally, if the codewords are indexed or labeled such that codewords which are close in Hamming distance are also close in the chosen distortion criteria (e.g., absolute distance) then the resulting encoding architecture will be relatively insensitive to transmission errors. This is because random bit errors in the channel will result in reproduction vectors which are "close" in the distortion criteria chosen by the codebook designer.

2.3.1 The FSCL Training Phase

In the *training phase*, the exemplar vectors in the winner-take-all layer are adjusted adaptively to statistically reflect the distribution of the training vectors. A training vector is applied to the input layer and compared to all of the exemplar vectors in the winner-take-all layer. Upon the completion of the comparisons, one node in the winner-take-all layer is selected to be the *winner* and the rest of the nodes are inhibited. The selection of the winner node depends on the product of the fairness function, $\mathcal{F}(u_i)$, and the distortion measure. The distortion measure of each input training vector is calculated with respect to the exemplar vector. *Common approaches for measuring the* distortion of the input and exemplar vectors are dot product, Euclidean distance, and absolute distance.

The exemplar vector of the winning node is adjusted, by an amount specified by the learning rate, so as to more closely represent the input vector. Finally, the winner node increments its update-counter,

u_i, and then the next training vector is presented to the network. Each node in the winner-take-all layer has a private update-counter, and the counters are used to influence the selection of the winner nodes. As a result, infrequently used exemplar vectors are adjusted, even if they had a larger distortion measure than other exemplar vectors [9].

When the learning phase is completed, all of the weight vectors in the winner-take-all layer are adapted, and the FSCL-derived codebook can be used in the encoder. As previously noted, the use of the modified distortion measure insures that the codewords are updated approximately the same number of times, thus maximizing the entropy of the codebook.

2.3.2 The Encoding and Decoding Phase

After the codebook has been constructed, input data vectors are coded by comparing each data vector with each of the codewords and then transmitting (or storing) the index of the (winning) codeword which yields the minimum distortion. Finding the minimum distortion codeword is a time-consuming task, but this operation is inherently parallel. The parallel hardware we have developed for the encoding is discussed in Section 3, and uses winner-take-all circuits in an Associative Memory encoder. Note that the winner-take-all circuits are used during training and encoding. However, only during training does adaptation of the codewords occur.

To decode the datum the receiver uses the received index to access a copy of the codebook to determine the reproduction vector used to represent the original data. This is a simple look-up process, and can be done without special hardware support in conventional RAM.

3 Vector Quantization as Associative Memory

VQ can be thought of as the process taking an input vector and matching it to the closest vector out of a set of vectors. Once the closest match is found, the index of the match is transmitted to the receiver. This process can also be viewed as an Associative Memory (AM), where the index is associated with a particular codeword.

As an example of an associative memory which uses Hamming distance and has very fast matching capabilities, consider the following design. The AM cell, shown in Figure 1, is a simple variant of the standard static RAM cell.
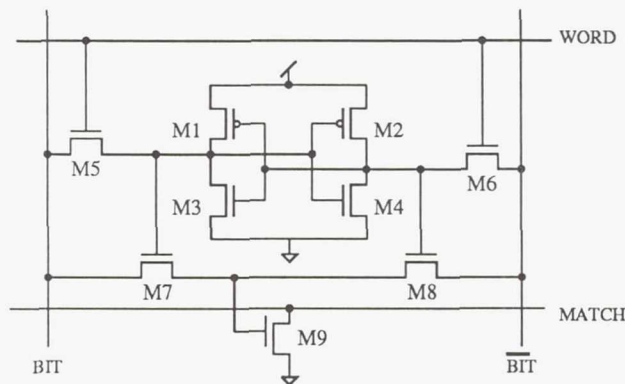
Figure 1: An Associative Memory cell. The transistors M7, M8 and M9 are the only additions to a standard static RAM. Their function is to draw the MATCH line low when the value on the BIT and $\overline{\text{BIT}}$ lines match the value stored in the static RAM core.

When in a matching mode, i.e. any time the WORD line is not asserted, if the value on the BIT line matches that stored by transistors M1 and M3, the transistor M7 becomes active and M9 is turned on and the MATCH line is drawn down. Similarly, if the $\overline{\text{BIT}}$ line matches the value stored in transistors M2 and M4, the transistor M8 becomes active and again M9 turns on.

The AM cells are arranged in an array such that the words to be matched all share the same MATCH line, as shown in Figure 2.
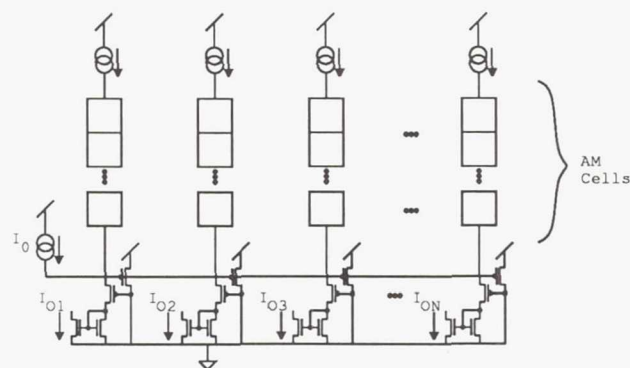

Figure 2: An array of associative memory cells, with their corresponding closest match circuitry. The circuitry which determines the closest match is a winner-take-all network which grows linearly with the number of words to match.

At the top of each column of AM cells is a current supply, which drives the MATCH line. In match mode as the value to be matched is shown to the column

via the BIT and $\overline{\text{BIT}}$ lines, the AM cells begin to sink current if the value stored in the AM matches that presented on the BIT and $\overline{\text{BIT}}$ lines. The circuitry at the bottom then determines which current coming in is the highest. The only I_{OUT} with any output current is the I_{OUT} for which the corresponding MATCH line that is the largest. This I_{OUT} is then used to gate the value of the given AM column to the output buffers.

4 Differential Vector Quantization

Differential Pulse Code Modulation (DPCM) can be used to perform quantization of image data in a CODEC (COder–DECoder) architecture. One example can be found in [10] which is based on DPCM but also utilizes a nonuniform quantizer and multilevel Huffman coding to reduce the data rate substantially below that achievable with straight DPCM. As a result of variable length coding, the compression ratio is different for different images depending on the statistics of that particular image. Furthermore, because of the different codeword lengths it is very difficult to detect and compensate for transmission errors even if an end-of-line reset is used to re-sychronize the encoder and the decoder.
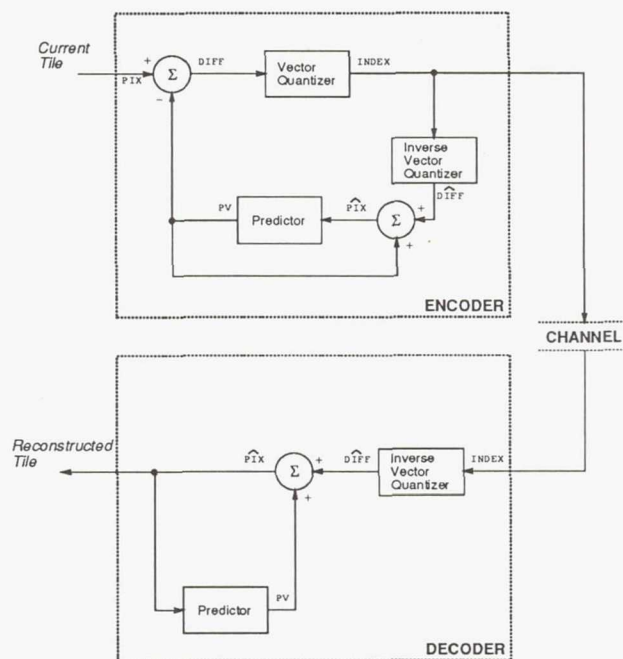

Figure 3: Block Diagram for Differential Vector Quantizer

Differential Vector Quantization (DVQ) (see Figure 3) incorporates the desirable qualities of both Differential Pulse Code Modulation (DPCM) and

VQ [11]. In DVQ, instead of scalar quantizing the scalar difference values (as in DPCM), vectors of difference values (difference tiles) are vector quantized and their codeword indices transmitted. At the receiver, the indices are used to look up the reconstruction difference vector which is then added to the predicted difference vector.

Because the VQ coder uses the FSCL derived codebooks, over a large sample of images the codewords are utilized relatively equally. Consequently, fixed length codes are used and the codec does not need to use synchronizing codes. Furthermore, as discussed earlier, by arranging the codewords so that codewords which are Hamming-close are also distortion-close we can achieve a significant level of error insensitivity, as the results in Section 5 clearly show.

5 Results

Table 1 shows the MSE obtained when DVQ codec using a FSCL codebook was used to compress eight images. For comparison, the MSE for a version of a DPCM codec, as described in [10], is also given. As can be seen, the DPCM codec yields MSEs which are lower than those of the DVQ codec. However, as can be seen in Figures 4 and 5 the images are virtually indistiguishable (these images are indistinguishable at full resolution, as well). Furthermore, as shown in Figure 6, the DVQ codec is significantly more robust to channel errors. In the left picture of Figure 6 line resynchronization was used in an attempt to minimize errors to one line of the image. Note that, even with resynchronization, the method results in truncated lines. Furthermore, when errors occur in the synchronizing codeword, lines are missed entirely.

6 Conclusions

We have presented an ANN based DVQ codec which is well suited for use in data compression applications such as Digital Video. There are three novel aspects of this work. First, Differential Vector Quantization is used to preserve edge features. Second, a new adaptive algorithm, known as Frequency-Sensitive Competitive Learning, is used to develop a vector quantizer codebook that eliminates the need for Huffman coding. Finally, in order to realize real time performance, a custom VLSI ASIC is being constructed to perform the associative memory functions needed in the vector quantization algorithm. The resulting codec exhibits greater compression and superior performance against channel bit errors than methods that use variable length codes.

7 References

[1] R. M. Gray, "Vector Quantization," *IEEE ASSP Magazine*, vol. 1, pp. 4–29, April 1984.

[2] J. Makhoul, S. Roucos, and H. Gish, "Vector Quantization in Speech Coding," *Proceedings of the IEEE*, vol. 73, pp. 1551–1588, November 1985.

[3] S. C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton, "Competitive Learning Algorithms for Vector Quantization," *Neural Networks*, vol. 3, pp. 277–290, 1990.

[4] A. K. Krishnamurthy, S. C. Ahalt, D. Melton, and P. Chen, "Neural Networks for Vector Quantization of Speech and Images," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 8, 1990.

[5] Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, vol. COM-28, pp. 84–95, January 1980.

[6] A. Buzo, A. H. Gray, Jr., R. M. Gray, and J. D. Markel, "Speech Coding based upon Vector Quantization," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, pp. 562–574, October 1980.

[7] S. P. Luttrell, "Derivation of a Class of Training Algorithms," *IEEE Transactions on Neural Networks*, vol. 1, pp. 229–232, June 1990.

[8] T.-C. Lee and A. M. Peterson, "Adaptive Vector Quantization Using A Self-Development Neural Network," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 8, 1990.

[9] P. Chen, "The Neural Shell: A Neural Networks Simulator," Master's thesis, The Ohio State University, March 1989.

[10] M. J. Shalkhauser and W. A. Whyte, Jr., "Digital CODEC for Real-Time Processing of Broadcast Quality Video Signals at 1.8 Bits/Pixel," Technical Report NASA TM-102325, National Aeronautics and Space Administration, Lewis Research Center, Cleveland, 1989. Prepared for Global Telecommunications Conference sponsored by IEEE, Dallas, Texas, Nov 27-30, 1989.

[11] C. W. Rutledge, "Vector DPCM: Vector Predictive Coding of Color Images," in *Proceedings of the IEEE Global Telecommunications Conference*, pp. 1158–1164, September 1986.

Table 1: MSE of NASA Codec and FSCL Differential Vector Quantization Algorithms

Picture	Bits Per Pixel (BPP)			MSE: No Errors			MSE: BER = 1/1000		
	codec	dvq 128	dvq 256	codec	dvq 128	dvq 256	codec	dvq 128	dvq 256
bird	2.56	1.75	2.00	6.5	20.2	14.5	2152	70.2	71.6
everest	2.17	1.75	2.00	5.0	16.8	12.6	981	44.1	34.3
fruity†	2.30	1.75	2.00	5.9	16.5	11.9	604	67.1	71.0
hall	2.97	1.75	2.00	7.7	17.8	12.7	1010	71.1	63.0
kitty†	2.75	1.75	2.00	5.9	14.0	10.1	964	65.7	47.9
lenna	2.37	1.75	2.00	5.3	11.4	8.3	1213	47.7	34.6
mandril	4.05	1.75	2.00	13.1	89.0	66.4	3107	305.7	201.1
planet†	2.09	1.75	2.00	4.6	13.9	10.0	1113	56.6	61.6
scene†	2.87	1.75	2.00	7.9	22.6	16.3	1993	75.7	75.7
sf†	2.89	1.75	2.00	9.5	57.1	41.9	3008	159.9	139.7

†These images were used in training the vector quantizer.
Error MSE's averaged over 3 trials.

Figure 4: Original image (left) and reconstructed image using the NASA codec algorithm (right)

Figure 5: Reconstructed image using differential vector quantization, FSCL with 128 codewords (left) and 256 codewords (right)

Figure 6: Reconstructed image with a noisy transmission channel (1 error per 1000 bits), NASA codec algorithm (left) and differential vector quantization using FSCL with 256 codewords (right)

Page intentionally left blank

GTEX: An Expert System for Diagnosing Faults in Satellite Ground Stations

*Richard Schlegelmilch, John Durkin
The University of Akron
Akron, Ohio

Edward Petrik
NASA Lewis Research Center
Cleveland, Ohio

Abstract

A research effort was undertaken to investigate how expert system technology could be applied to a satellite communications system. The focus of the expert system is the satellite ground stations. Diagnostic procedures associated with the ground stations are very demanding. Knowledge about the operational characteristics, communications strategies and the associated electronics of the ground station is required.

A proof of concept expert system called GTEX (Ground Terminal Expert) was developed at The University of Akron in collaboration with NASA Lewis Research Center. The objective of GTEX is to aid in diagnosing data faults occurring with a digital ground terminal. Thou, research focused on large systems, this strategy can also be applied to the Very Small Aperture Terminal(VSAT) technology. An expert system which detects and diagnoses faults would enhance the performance of the VSAT by improving reliability and reducing maintenance time.

GTEX is capable of detecting faults, isolating the cause and recommending appropriate actions. Isolation of faults is completed to board-level modules. A graphical user interface provides control and a medium where data can be requested and cryptic information logically displayed. Interaction with GTEX consists of user responses and input from data files. The use of data files provides a method of simulating dynamic interaction between the digital ground terminal and the expert system. GTEX as described is capable of both improving reliability and reducing the time required for necessary maintenance.

GTEX was developed on a personal computer using the Automated Reasoning Tool for Information Management (ART-IM) developed by Inference Corporation. Developed for Phase II digital ground terminal, GTEX is a part of the Systems Integration Test and Evaluation(SITE) facility located at NASA Lewis Research Center.

1.0 Introduction

System studies performed during the late 1970's indicated that advanced communications satellite technologies should be developed to utilize the Ka-band (30/20 GHz) spectrum. In order to demonstrate Ka-band satellite communications systems, NASA Lewis Research Center has conducted an advanced space communications program. This program will meet the needs of future NASA missions and will infuse advanced technologies into the commercial sector. One objective of the program is to apply advanced digital logic to space communications including the satellite ground terminal and network control. The program focuses on several major areas: advanced modulation and coding; space based processing and control; and ground based processing and control. The goal of ground based processing and control focus is to develop cost efficient terminals. Expert systems are being applied to diagnose ground terminal failures and provide autonomous operation.

2.0 Background

Extensive development of satellite communications is currently under way at the NASA Lewis Research Center. Using proof-of-concept subsystems and components, (IF switch matrices, solid-state amplifiers, traveling-wave tube, high-power amplifiers and low-noise receivers) a Ka-band satellite communications network simulation known as the Satellite Integration, Test and Evaluation (SITE) facility has been developed. This facility allows modulated data to be used to characterize the effect of microwave components on Bit Error Rate (BER) performance and SITE supports voice, data, and video through a Time Division Multiple Access (TDMA) burst terminal and a three ground terminal network.

The ground terminal is a major element of the simulator and one of the more complex. Each ground terminal must be capable of acquiring satellite and network timing, maintaining synchronization to the network, and transmitting and receiving bursted data from other ground terminals in the network.

Each ground terminal, shown in Figure 1, contains a 221.184-MHz system clock, transmission and reception timing and control circuits, compression or expansion first-in, first-out memories (FIFO) separate user clocks and their associated control circuits, an orderwire processor microcomputer, a user interface controller, and a 221.184-MHz serial minimum-shift key(SMSK) burst modulator/demodulator (modem) (Ivanic, et al. 1989).

In the SITE ground terminal, users are simulated by a bit-error-rate test set consisting of a data generator (transmitting user) and a data checker (receiving user). A controlling computer creates realistic traffic patterns with users of varying data rates entering and leaving the system. The raido frequency (RF) components and links of the communication system can degrade the user's data which often results in bit errors. The users need to know the degree of data degradation to determine whether to tolerate it or compensate for it. Data degradation can be quantified by using a bit-error-rate (BER) figure. A BER provides a performance measure of the RF communication links, the RF and digital subsystems, and the overall satellite communication system.
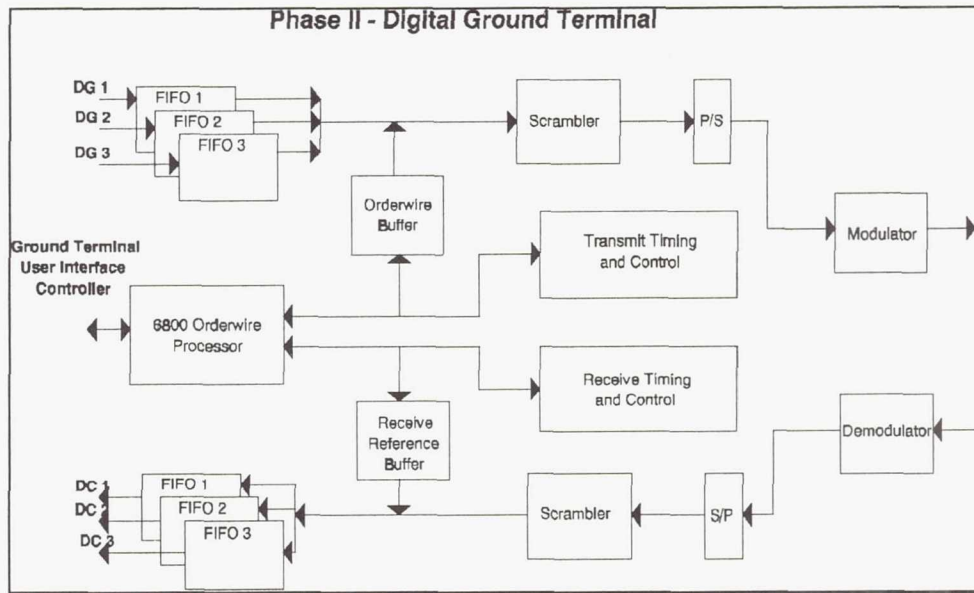
Figure 1. - SITE Digital Ground Terminal Block Diagram

Providing an interface to the users, the SITE ground terminals control the timing and the paths of the users' data transmission. Presently, each of the SITE ground terminals is capable of interfacing to three transmitting and three receiving users. The transmitting ground terminal multiplexes the data it receives from each transmitting user and bursts the data to the satellite at a high data rate. The receiving ground terminal receives the high-bursts and demultiplexes them so that each receiving user gets the proper data. (Shalkhauser 1988)

Since the SITE digital ground terminal is itself a prototype, new errors frequently occur. This required the scope of GTEX to be very concentrated. The errors associated with the transmitting and receiving of user data is the primary focus of GTEX.

Being developed as a diagnostic aid, GTEX will be used as a demonstrational prototype showing the feasibility of applying expert system technology in the area of high-rate digital communications at NASA Lewis Research Center.

4.0 Development Environment

GTEX is being developed using the Automated Reasoning Tool for Information Management (ART-IM) by Inference Corporation. ART-IM is a C-based toolkit for the development of rule-based, or knowledge-based, expert systems (ART-IM 1991). GTEX was developed on a low-cost development and deployment environment a personal computer(PC) running MS-DOS.

ART-IM supports three methods of programming styles; procedural, ruled-based and object-oriented. The procedural language supported by ART-IM provides basic function calls, and allows simple interactions and conditionals to be performed. The rule-based structure uses the rule as the fundamental unit. Reacting to changes in the working memory, the rule can then fire

105

or execute based on the dynamic order of the changes which occur. Objects in ART-IM are represented by a schema. Control of an object is managed by sending a message to that object. An object reacts to a message by searching itself for an appropriate method and executing the actions associated with that method.

The ART-IM procedural language can be extended using the 'C' programming language. User functions can be written in C and included with the ART-IM program which can be used like any other ART-IM function. Since functions defined in C are complied versus interpreted, the result is faster execution. This capability of ART-IM was fundamental in the development of the user-interface discussed in this paper.
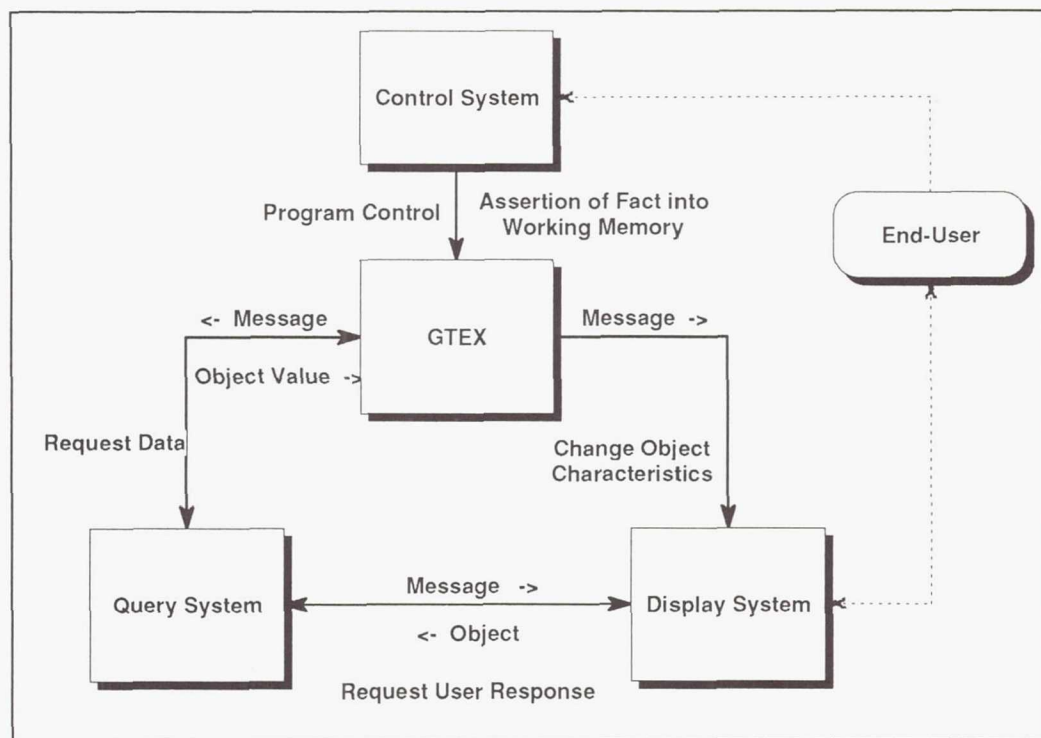
5.0 System Overview

Figure 2. - GTEX Architecture

The GTEX architecture is divided into the following subsystems:

- Display subsystem
- Control subsystem
- Query subsystem
- Knowledge base

A modular approach was beneficial in system development. An extended ART-IM function was constructed for communications between the knowledge base, display and query subsystems. This implementation allows communications to be established from both ART-IM and C. Syntax would be as follows:

calling from ART-IM:

(send-message *system* "message")

calling from C:

aofnSendMessage(a_art_symbol(sSystemType), a_art_string(sMessage));

Communications with the knowledge base are achieved by asserting a fact into working memory corresponding to the end-users action. Messages each subsystem handles will be described in detail in the designated subsection.

The block diagram of the GTEX system architecture is shown in Figure 2. The diagram shows each subsystem and the possible communications paths.

5.1 Display Subsystem

A graphical environment was chosen to provide a dynamic interface. The GTEX interface shown in Figure 3 is capable of displaying dynamic system block diagrams, system status, and provide interactive dialog with the end-user. Dynamic system diagrams alert the end-user to changing conditions of the simulated digital ground terminal operation. Detailed diagrams of individual subsystems within the SITE ground terminal can also be displayed. Diagnostic messages regarding the inferencing process are conveyed to the end-user. Interactive dialog, implemented as a pop-up window, is provided to query information and provide a medium to display important messages.

The display subsystem receives a message string from either the knowledge base or query subsystem. The message string is in the form:

"Message-Type Message-Action"

Table 1, enumerates the message string possibilities.

Message Type	Message Action
Display	Page-Title
Update	Object
Initialize	-
Warning	Warning Message
Message	Message Number
Dialog	Object & Property

Table 1 - Message strings for Display Subsystem

107

Phase II - Digital Ground Terminal

DG 1
DG 2
DG 3

Ground Terminal
User Interface
Controller

Modulator

Demodulator

DC 1
DC 2
DC 3

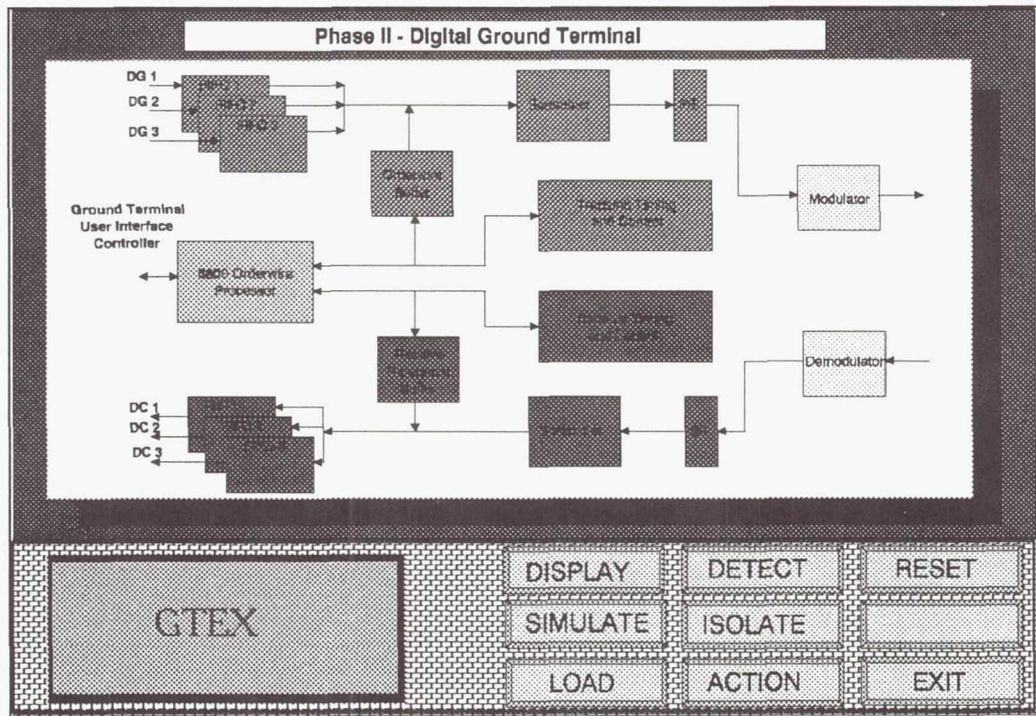| DISPLAY | DETECT | RESET |
| SIMULATE | ISOLATE | |
| LOAD | ACTION | EXIT |

GTEX

Figure 3 - GTEX Interface

The interface to GTEX was developed in C using the Essential Graphics Library developed by South Mountain Software. The modularity of the system was preserved by only employing 'C' constructs in development. A pseudo object-oriented architecture was constructed to simplify development of required displays.

5.2 Control Subsystem

Since the diagnostic procedures are being performed in cooperation with a technician, control of GTEX is necessary. Basic actions such as starting the diagnostic process, resetting the system and selecting the type of system dialog are originated by the control system.

Control of GTEX is performed by a function whose execution is controlled by ART-IM. The function, known as an asynchronous function, is called before the first rule-firing, between rule-firings and after the last rule-firing.

Developed using the Essential Graphics library, the control system responds to mouse interaction. Once interaction has been detected, validation tests are performed and proper procedures taken. The final result is the assertion of the corresponding fact into working memory. Control is immediately returned to the knowledge base if no interaction has occurred.

5.3 Query Subsystem

The end-user of GTEX has an option of two data input formats. A data file, simulating dynamic interaction with the SITE ground terminal is preferred. If data cannot be obtained from the data file, or at user request, GTEX will prompt the user for the required data.

The simulated data file is a ASCII text file with the following record format:

Object.Object-Property = Object-Value

Initilialization of the simulator is controlled by a message received from the knowledge base. The message received contains the name of the ASCII data file. The ASCII file is parsed and the corresponding object and properties are created and values inserted. Since the simulator objects mimic the objects associated with the knowledge base, the ART-IM schema notation is used.

To query a value, a message is received from the knowledge base denoting an object and the corresponding property of the value required. The inquiry begins by checking a status flag to decide if data is to be retrieved from the simulator. Found to be true, an object-property match is performed and corresponding data returned. Dialog is conducted with the end-user either when, a value is unavailable from the simulator or the status flag indicates that the end-user to be queried. During the dialog, the system halts execution until valid data has been obtained from the end-user.

5.4 Knowledge Base

GTEX performs the following tasks:

- Fault Detection
- Fault Isolation
- Fault Recovery Recommendation

The diagnostic procedure begins by determining the initial system configuration. Information is gathered about each data channel regarding the transmit, receive channels, and data rate. General information about the SITE ground terminal is also obtained. BER measurements are performed and evaluated on each of the user data channels to determine if tolerance limits have been exceeded. Once tolerance limits have been exceeded, the end-user is informed of the discrepancy and isolation of the fault begins.

Three levels of isolation are performed. The initial stage of isolation determines which side, transmit or receive, of the user channel is causing the error. Isolation is then performed to determine which of the corresponding subsystems is in error. Finally the corresponding circuit board is isolated. Once the fault has been isolated and end-user informed, GTEX can be requested to recommend a fault recovery recommendation

6.0 Knowledge Base Architecture

The knowledge base is divided into the following rule categories:

- Display rules
- Diagnostic rules
- Demon rules

Each rule category has its own salience or relative priority which is used in scheduling the rule for firing. This technique was implemented since the version of ART-IM used for development was unable to handle rule sets. Table 2 lists the salience values assigned to the individual categories.

Rule category	Salience
Display Rules	200
Demon rules	100
Diagnostic rules	0

Table 2 - Table of Rule Salience

6.1 Diagnostic Rules

The diagnostic rules are the foundation of the knowledge base. These rules perform the fundamental steps required in diagnosing the digital ground terminal. They are responsible for asserting facts into the working memory, which results in the execution of the detailed procedures, implemented using "demon rules" discussed in this paper, for the current level of diagnostics.

The diagnostic rules have the lowest priority of any rule group. This priority level is important in maintaining system integrity. The integrity is preserved by insuring that the diagnostic rules are reacting to current data. An example of a diagnostic rule is shown.

```
(DEFRULE Determine-Channel-Configurations
        (SCHEMA Ground-Terminal-Configuration
            (Attenuators-Connected YES|NO)
            (RF-Transponder-Connected NO))
=>
        (ASSERT (Determine Channel 1 Configuration))
        (ASSERT (Determine Channel 2 Configuration))
        (ASSERT (Determine Channel 3 Configuration)))
```

This rule assumes that the ground terminal configuration was perviously determined. The connection of attenuators and the RF-transponder with the digital ground terminal is checked. If the attenuator connection is known and the RF transponder found not connected, then facts regarding the user data channel configuration are asserted in working memory.

110

6.2 Demon Rules

Sitting dormant in the knowledge base are demon rules. The responsibility of the demon rules are to perform detailed diagnostic procedures. They are activated by the diagnostic rules which assert appropriate facts in working memory. The modification of objects is the primary procedure performed.

An example of a demon rule is shown.

```
(DEFRULE Determine-Channel-Configuration-Demon
      (DECLARE (SALIENCE ?*demon-salience*))
      (SCHEMA ?Channel
            (INSTANCE-OF User-Channel)
            (Channel-Number ?Number))
      ?demon <- (Determine Channel ?Number Configuration)
=>
      (MODIFY (SCHEMA ?Channel
            (Transmit-Channel
                  =(SEND TX-Channel-Number User-Interface-Computer ?Number))
            (Receive-Channel
                  =(SEND RX-Channel-Number User-Interface-Computer ?Number))
            (Data-Rate
                  =(SEND Data-Rate User-Interface-Computer ?Number))))
      (RETRACT ?demon))
```

For each instance of a user channel in working memory, the rule fires which modifies the object after obtaining the data desired. The data in this case is the transmit and receive channels associated with this data channel and the corresponding data rate. Actual data values are obtained by sending a message to the appropriate object.

6.3 Display Rules

Display rules are responsible for setting and updating the display based on schema values in the knowledge base. They are also responsible for intercepting and processing fact messages from the control system. Rules regarding display have the highest priority to ensure that the display is always updated. An example of a display rule is shown below.

```
(DEFRULE Select-User-Dialog
      (DECLARE (SALIENCE ?*display-salience*))
      ?display-fact <- (ACTION SELECT USER-DIALOG)
=>
      (BIND ?*simulator* (send-message display-system "dialog simulator"))
      (send-message display-system "message 102")
      (retract ?display-fact))
```

This rule shows a fact message from the control system which informs the knowledge base that the end-user wants to select the type of user-dialog. The right-hand side of the rule then sends two messages to the display system.

111

7.0 Summary

As more emphasis is placed on developing low-cost efficient ground terminals, expert system technology will play an essential role. GTEX is being developed as a prototype, demonstrates the ability of expert systems to provide needed assistance in ground terminal operation. The modular approach taken will allow GTEX to adapt to new system designs. Adhering to this viewpoint, future consideration includes modifying GTEX to enhance the capabilities of the High-Burst Rate Link Evaluation Terminal (HBR-LET) designed for the Advanced Communications Technology Satellite (ACTS) Project at the NASA Lewis Research Center.

8.0 Acknowledgments

9.0 References

Ivanic, William D., Andro, Monty, Nagy, Lawarence A., Budinger, James M., Shalkhauser, Mary Jo, "Satellite-Matrix-Switched, Time-Division-Multiple-Access Network Simulator", NASA Technical Paper 2944, October 1989

Shalkhauser, Mary Jo, "Satellite Ground-Terminal User Simulation", NASA Technical Memorandum 100234, January 1988

"ART-IM Programming Language Reference", Inference Corporation, El Segundo CA. 90245, 1991

"ART-IM in the DOS Environment", Inference Corporation, El Segundo CA. 90245, 1991

* Principle Contact : Richard Schlegelmilch, Electrical Engineering, The University of Akron, Akron Ohio, 44325.

LABORATORY MEASUREMENTS OF ON-BOARD SUBSYSTEMS

P.P. NUSPL, G. DONG AND H.C. SERAN
INTELSAT
WASHINGTON, D.C., 20008-3098

SUMMARY

GOOD PROGRESS HAS BEEN ACHIEVED ON THE TEST BED FOR ON-BOARD SUBSYSTEMS FOR FUTURE SATELLITES. THE TEST BED IS FOR SUBSYSTEMS DEVELOPED UNDER PREVIOUS INTELSAT R&D CONTRACTS. FOUR TEST SETUPS HAVE BEEN CONFIGURED IN THE INTELSAT TECHNICAL LABORATORIES:

1. TDMA ON-BOARD MODEM (MODEM, MELCO);
2. MULTICARRIER DEMULTIPLEXER DEMODULATOR (MCDD, TELESPAZIO/ALCATEL);
3. IBS/IDR BASEBAND PROCESSOR (BBP, NEC); AND
4. BASEBAND SWITCH MATRIX (BSM, NEC).

THE FIRST THREE SERIES OF TESTS ARE COMPLETED AND THE TESTS ON THE BSM ARE IN PROGRESS. DESCRIPTIONS OF TEST SETUPS AND MAJOR TEST RESULTS ARE INCLUDED IN THIS POSTER PRESENTATION; THE FORMAT OF THE POSTER IS OUTLINED BELOW. A COMPANION PAPER DISCUSSES THE SYSTEMS BENEFITS AND CONSTRAINTS, AND SUMMARIZES THE STATUS OF THESE ON-BOARD TECHNOLOGIES [REF. 1].

TITLE PAGE

```
*********************************************** --------------------------
*                                            * -                        -
*             *                *             * -            -           -
*             MODEM            *             * -           MCDD         -
*             *                *             * -            -           -
*                                            * -                        -
*********************************************** --------------------------
+++++++++++++++++++++++++++++++++ ...........................................
+                               + :                    :  BBP        :     :
+             +                 + :.......................................:
+             BSM               +++++++++++++++++++++++++++++++++++++++++++++
+             +                 +             +            +           +
+                                                                      +
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

*THIS PAPER WAS PRESENTED IN SESSION 4: CONCURRENT POSTER PRESENTATIONS AND DEMONSTRATIONS. ROW AND COLUMN NUMBERS EXPLAIN THE POSTER LAYOUT.
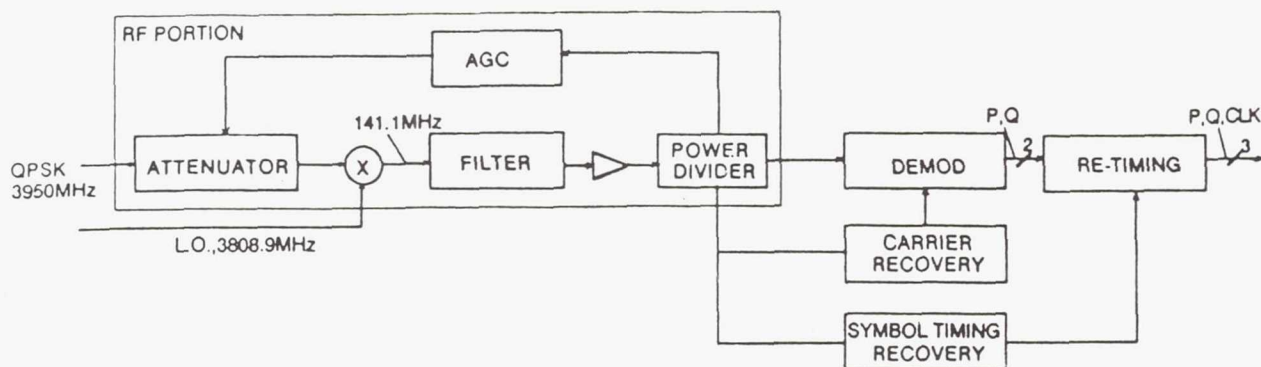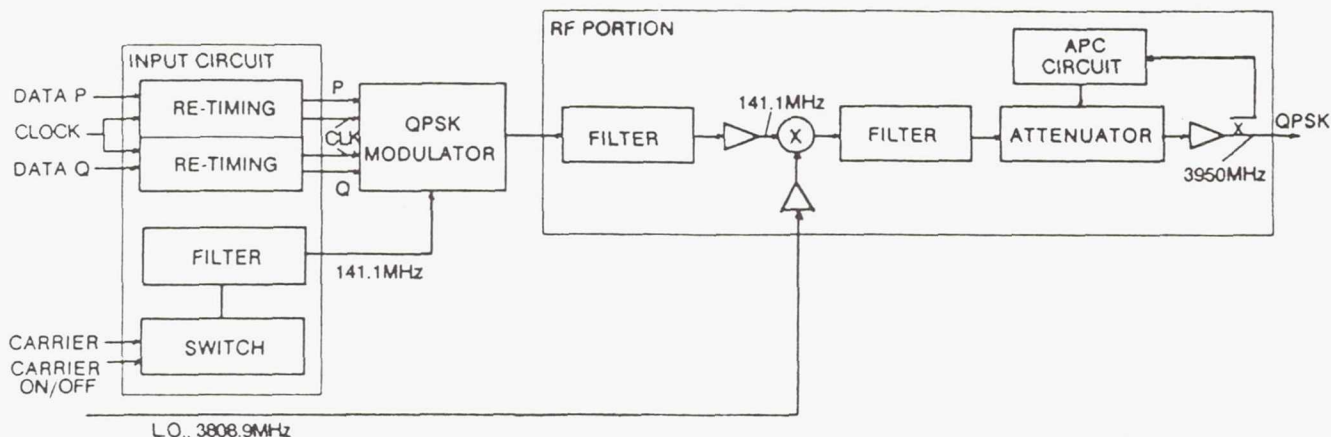
FIGURE 1 (A)   ON-BOARD DEMODULATOR

FIGURE 1 (B)   ON-BOARD MODULATOR

THE CODE GENERATOR PRODUCES
THE P, Q, CLOCK AND GATE SIGNALS (IN BURST MODE).

THE CODE ERROR DETECTOR ACCEPTS THE P, Q, AND CLOCK SIGNALS
AND THE STATUS AND APERTURE SIGNALS FROM THE CODE GENERATOR
AND MAKES BER, UNIQUE WORD MISSING DETECTION (UWMD) TESTS ETC.

THE ADAPTER IS AN INTERFACE BETWEEN THE CODE GENERATOR,
CODE ERROR DETECTOR AND THE EARTH STATION (E/S) MODEM.

ATTENUATOR 1 IS USED TO ADJUST THE INPUT SIGNAL LEVEL
TO THE UPCONVERTER, WHICH IS ABOUT -20 DBM.

SWITCH 1 CONTROLS THE SIGNAL CONNECTION FOR SIGNAL MONITORING
AND E_b/N_o CALIBRATION DURING THE MEASUREMENT PROCESS.

ROW 1, COLUMN 1

THE BURST 3950-MHz RF SIGNAL IS FED TO THE RF PORTION WHICH INCLUDES
THE DOWNCONVERTER, THE IF ROLL-OFF FILTER AND THE AGC CIRCUIT.

THE DEMODULATION CIRCUIT IS A COHERENT DETECTOR.

THE CARRIER RECOVERY CIRCUIT CONSISTS OF A TIMES-FOUR MULTIPLIER,
A TANK-LIMITER WITH AFC (AUTOMATIC FREQUENCY CONTROL)
AND A DIVIDE-BY-FOUR CIRCUIT.

THE SYMBOL-TIMING-RECOVERY CIRCUIT CONSISTS OF THE IF SQUARING CIRCUIT
AND TANK-LIMITERS.

IN THE MODULATOR, THE P AND Q STREAMS AND THEIR CLOCK OF 60.416 MHz
(DATA RATE IS 120.832 MBIT/S) ARE RECEIVED BY THE RETIMING CIRCUIT.
P AND Q STREAMS ARE SYNCHRONIZED BY THE CLOCK.

THIS SWITCH IS CONTROLLED BY THE CARRIER ON-OFF SIGNAL
FROM THE TEST SET AND IT CONTROLS THE OUTPUT OF THE MODULATOR.
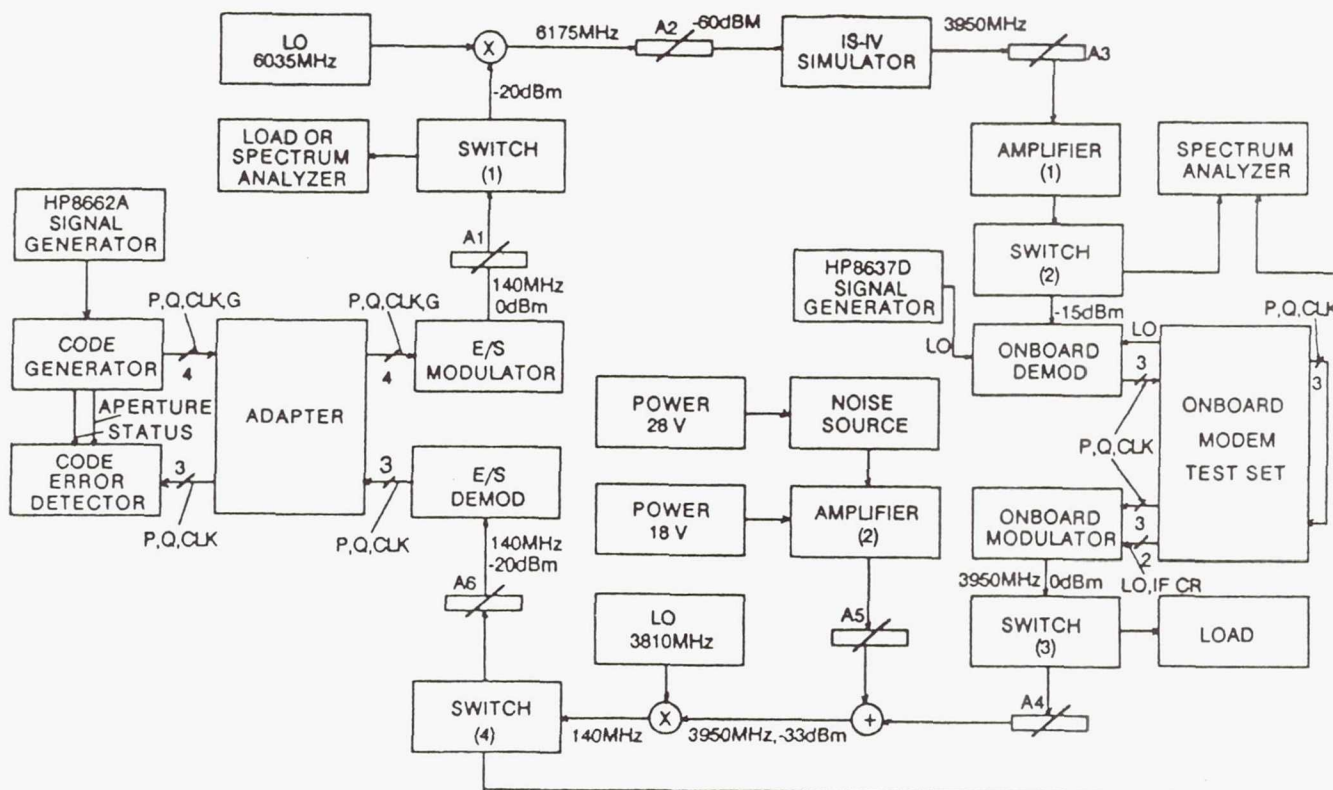
FIGURE 2   BER TEST SETUP

ROW 1, COLUMN 2

The equipment from the Code Generator to the Upconverter simulates the function of the E/S transmitter in which the QPSK-modulated 6-GHz signal is produced.

The Low Noise Amplifier (LNA) is in the INTELSAT-IVA Transponder Simulator. Attenuator 2 is used to control the signal level to the LNA amplifier and to calibrate the uplink E_b/N_o.

A Downconverter with LO frequency of 2225 MHz is also in the INTELSAT-IVA Transponder Simulator.

Switch 2 is used to control the signal connection for signal level and spectrum measurement or monitoring in the test process; this is done without changing the physical connection in order to improve the measurement accuracy and facilitate operation.

The demodulated P and Q signals and recovered Clock from the on-board Demodulator pass through the buffer and conditioning circuits in the Modem Test Set.

Switch 3 is for downlink E_b/N_o calibration and signal monitoring. The Variable Attenuator 5 is used to control the noise level to reach the required downlink E_b/N_o. Switch 4 is used for downlink E_b/N_o calibration and spectrum monitoring without changing the physical connection.

In the BER test, the measured performance is the summation of the uplink and downlink BERs. If the uplink (or downlink) E_b/N_o is very high, for instance over 40 dB, the measured result will mainly indicate the performance of the downlink (or uplink respectively).

The E_b/N_o calibrations are different for the uplink and the downlink. For the uplink, the noise is mainly the thermal contribution from the LNA in the INTELSAT-IVA Transponder Simulator. The Spectrum Analyzer is used to measure the signal level (unmodulated carrier) and the noise power density (in dBm/Hz).

For the downlink, an independent noise source is used. The Spectrum Analyzer measures the unmodulated carrier level under very weak noise conditions, and the noise spectral power density without the signal. In the entire calibration process the signal connection is controlled by switches without changing any physical connection, so the calibration error is reduced significantly.

ROW 1, COLUMN 3

THE DEMULTIPLEXER SEPARATES THE CHANNELS USING A PER-CHANNEL, ANALYTIC SIGNAL APPROACH.

THE DEMODULATOR IS A SINGLE-CHANNEL DEMODULATOR THAT RECOVERS THE TRANSMITTED BIT STREAM AND OUTPUTS IT TO A BASEBAND SWITCH MATRIX. THE BIT RATE OF THIS MCDD CANNOT BE VARIED AND ONLY ONE CHANNEL CAN BE PROCESSED AT ANY ONE TIME. THE INPUT FDMA SIGNAL HAS A 10-MHz BANDWIDTH AND CONSISTS OF 3 CHANNELS AT 4.4 MBIT/S TRANSMISSION RATE, OR 12 CHANNELS AT 1.1 MBIT/S TRANSMISSION RATE; IT IS SAMPLED AT A RATE OF ABOUT 20 MHz.

AN ANALOG INPUT INTERFACE IS PROVIDED THAT IS ABLE TO ACCEPT THE SIGNAL AT INTERMEDIATE FREQUENCY (140 MHz), TO PERFORM THE ANTI-ALIASING FILTERING AND THE DOWN-CONVERSION TO BASEBAND, SO THAT THE FINAL ANALOG-TO-DIGITAL CONVERSION IS DONE AT NYQUIST RATE.

AT THE OUTPUT OF THE MCDD, A DIGITAL-TO-ANALOG CONVERTER IS USED FOR THE PURPOSE OF TESTING, AND ALLOWS AN OSCILLOSCOPE TO BE USED TO OBSERVE SIGNAL CONSTELLATIONS AND OTHER SIGNIFICANT PARAMETERS.

THE FDMA SIGNAL GENERATOR CONSISTS OF A BANK (THREE IN THIS CASE) OF MODULATORS WHICH HAVE THE SAME CONFIGURATIONS BUT THEIR CARRIER FREQUENCIES CAN BE SELECTED INDEPENDENTLY WITHIN CERTAIN RANGES.

THE HP3326A SYNTHESIZER PROVIDES THE REQUIRED CLOCK SIGNAL TO THE HP3762A DATA GENERATOR WHICH PRODUCES THE DATA SEQUENCE AND CLOCK.

THE ALCATEL SYNTHESIZER PROVIDES THE SOURCE FREQUENCY TO THE FDMA SIGNAL GENERATOR. IT CAN PRODUCE A MAXIMUM OF 3 MODULATED CARRIERS SIMULTANEOUSLY. THE HP3708A NOISE TEST SET IS USED TO INTRODUCE GAUSSIAN NOISE IN THE CHANNEL UNDER TEST.

AO IS THE OUTPUT ATTENUATOR INSIDE THE FDMA SIGNAL GENERATOR. ATTENUATORS A1 AND A2 ARE USED TO ADJUST THE SIGNAL LEVEL TO MEET THE REQUIREMENTS OF THE HP3708A NOISE TEST SET AND THE MCDD. THE HP8566B IS FOR SPECTRUM MONITORING AND ANALYSIS.
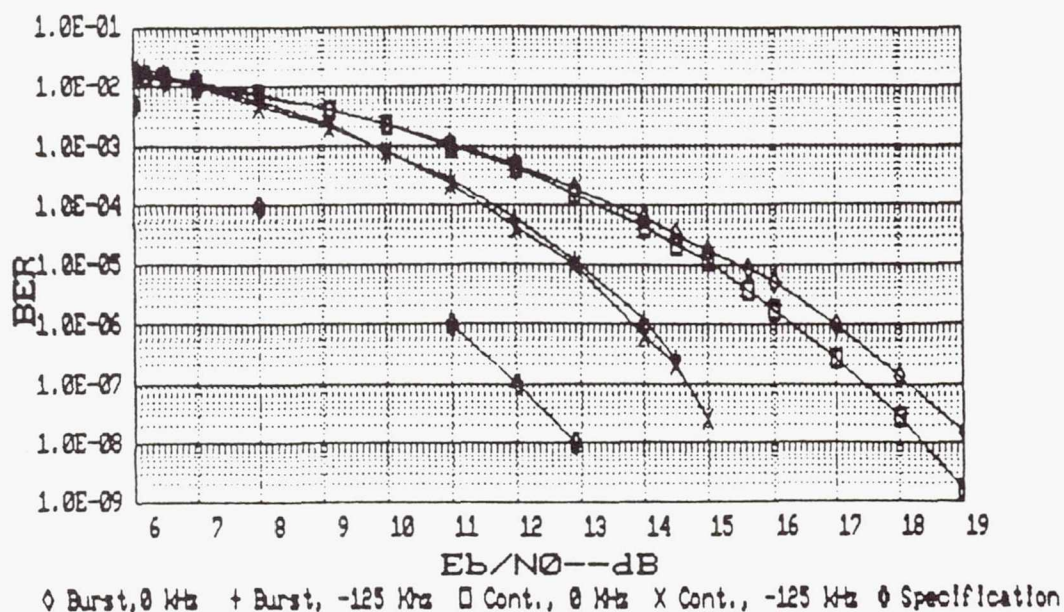
ROW 1, COLUMN 4

FIGURE 3   UPLINK BER FOR MODEM

THE UPLINK BER IS AN INDICATION OF THE ON-BOARD DEMODULATOR
PERFORMANCE. THE BER VERSUS E_b/N_o FOR THE BURST MODE AND CONTINUOUS
MODE ARE SIMILAR.   THE UPLINK E_b/N_o NEEDS TO INCREASE BY 2.1 TO 4.7 DB
IN ORDER TO GET THE SAME BER AS DEFINED IN THE SPECIFICATION.   THE
DEGRADATION REDUCES TO 0.7 TO 1.6 DB WHEN THE LO FREQUENCY OFFSET IS
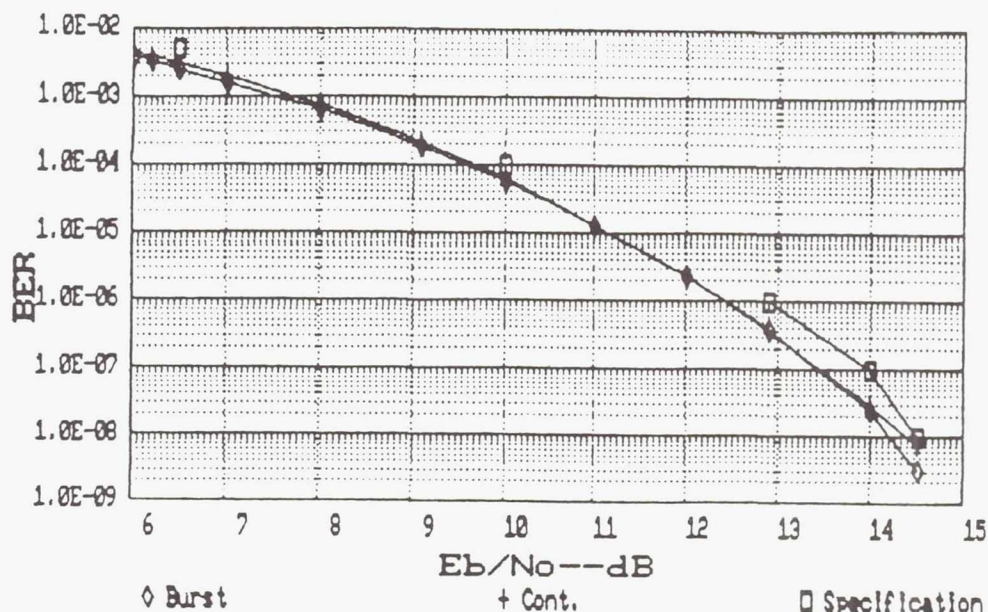AT ABOUT -125 KHZ OR THE CARRIER FREQUENCY OFFSET IS +125 KHZ.

ROW 2, COLUMN 1

FIGURE 4   DOWNLINK BER FOR MODEM

THE DOWNLINK BER REFLECTS THE ON-BOARD MODULATOR AND THE E/S
DEMODULATOR PERFORMANCE.   THE BER VERSUS E_b/N_o FOR THE BURST MODE
AND THE CONTINUOUS MODE ARE VERY SIMILAR.   FOR THE BURST MODE,
THE MEASURED BER VERSUS E_b/N_o IS BETTER THAN THE SPECIFICATIONS,
BY ABOUT 0.3 TO 0.6 DB FOR THE ON-BOARD MODEM
AND BETTER BY ABOUT 1.1 TO 2.0 DB FOR THE E/S MODEM.

DURING TESTS OF AMPLITUDE VARIATIONS, THE SWITCH ON THE CODE GENERATOR
WAS USED TO SELECT THE OPERATION MODE AND MEASURE THE CARRIER LEVEL
WITH THE SPECTRUM ANALYZER.   IN THE FIXED MODE (NO MODULATION) BOTH P
AND Q CHANNELS CAN BE SET WITH SWITCHES FOR A CONSTANT 0 OR 1, WHICH
PRODUCES AN AMPLITUDE VARIATION OF 0 DB.   IN THE CONTINUOUS MODE A
PSEUDO-RANDOM SEQUENCE IS SELECTED TO MODULATE A CONTINUOUS CARRIER.
THE PEAK-TO-PEAK AMPLITUDE VARIATION IN THIS CASE IS 0.6 DB,
SO THE MODEM MEETS THE SPECIFICATION OF ±0.5 DB AMPLITUDE VARIATION.

CARRIER ON/OFF ISOLATION WAS MEASURED AS 55 DB
AND EXCEEDS THE SPECIFIED 50 DB.

THE SPECIFICATION OF PROBABILITY OF UNIQUE WORD MISSED DETECTION (UWMD)
IS BETTER THAN 1×10^{-8} WHEN THE E_b/N_o IS EQUAL TO 7 DB.
THE TEST RESULT SHOWS THAT ABOUT 2 DB E_b/N_o INCREASE IS NEEDED
TO MEET THAT UWMD PROBABILITY.

ROW 2, COLUMN 2
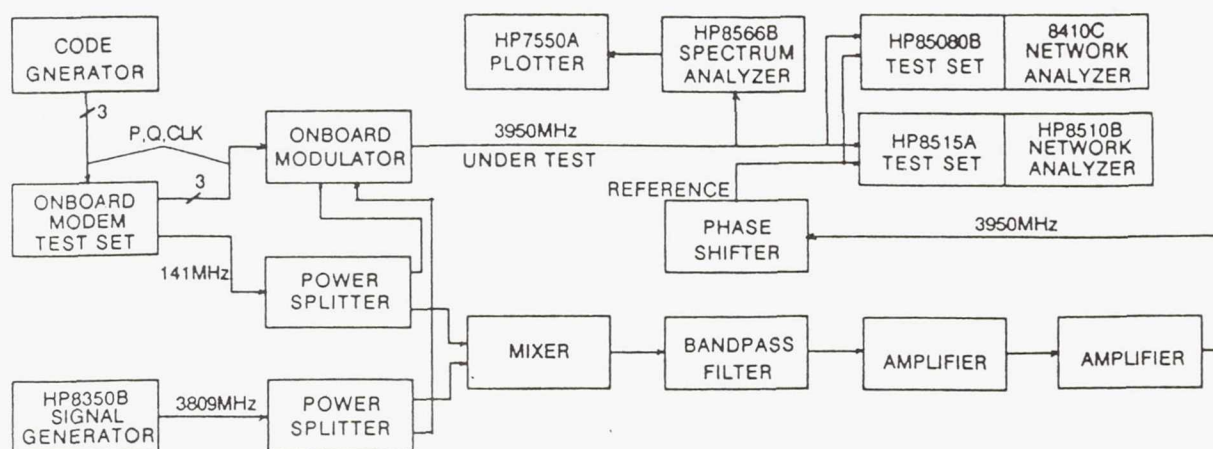
FIGURE 5   TEST SETUP FOR CARRIER PHASE SHIFT
AND AMPLITUDE VARIATION

THE HP8510B NETWORK ANALYZER IS USED FOR PHASE SHIFT MEASUREMENTS BY
COMPARING THE MODULATED CARRIER WITH THE REFERENCE CARRIER.

SINCE THE PURPOSE IS TO MEASURE THE RELATIVE PHASE DIFFERENCES AMONG
THE 00, 01, 11, 10 PHASES, THE ACCURACY OF THE ABSOLUTE VALUE IS NOT
VERY IMPORTANT.

WITHIN THE ON-BOARD MODULATOR, THE P AND Q DATA MODULATE THE IF CARRIER
OF 141 MHz FIRST, THEN THE IF CARRIER IS UPCONVERTED TO 3950 MHz. THIS
MODULATED SIGNAL IS SENT TO THE NETWORK ANALYZER FOR PHASE SHIFT
MEASUREMENTS.

THE REFERENCE SIGNAL HAS THE SAME RF FREQUENCY AND IS PHASE COHERENT TO
THE MODULATED CARRIER (UNDER TEST) OF THE ON-BOARD MODULATOR.

THE OUTPUT OF THE MIXER HAS THREE PRODUCTS AT VERY CLOSE LEVELS AT 3950
MHz (SUMMATION), 3809 MHz (LO) AND 3668 MHz (DIFFERENCE) RESPECTIVELY.
THE FREQUENCY OF THE DESIRED PRODUCT IS 3950 MHz AND THE FILTER SELECTS
THE DESIRED ONE.

THE PHASE SHIFTER IN THE REFERENCE CHANNEL IS USED TO COMPENSATE THE
STATIC PHASE DIFFERENCE BETWEEN THE TWO CHANNELS.

TEST RESULTS ARE 0.0, 90.7, 180.2, 269.4 DEGREES FOR THE FOUR PHASE
STATES AND MEET THE SPECIFICATIONS (0, 90, 180, 270, ±2 DEGREES).
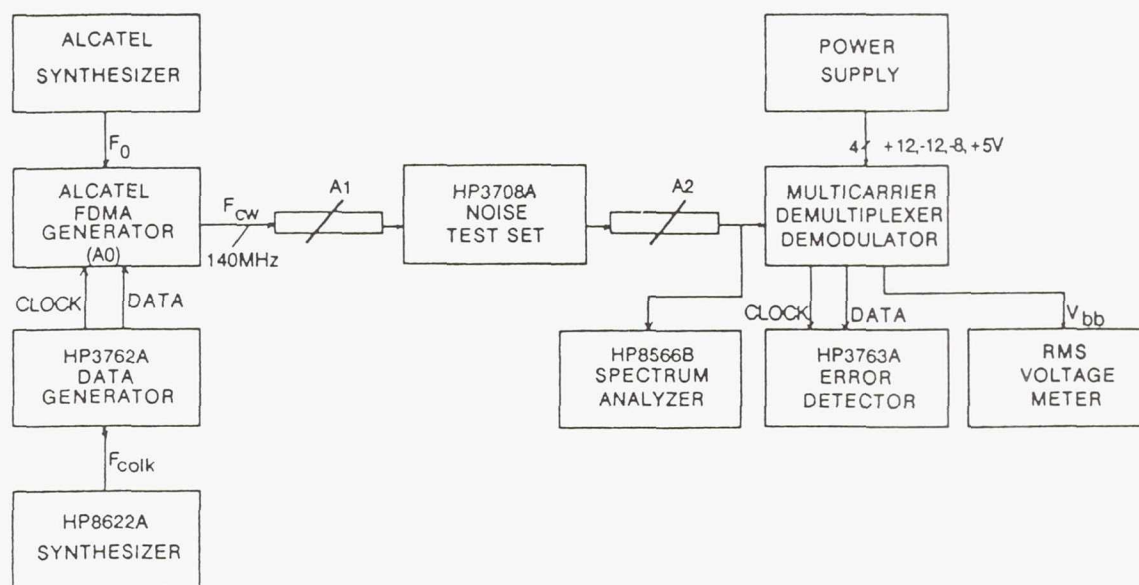
ROW 2, COLUMN 3

FIGURE 6   BER TEST BED FOR MCDD

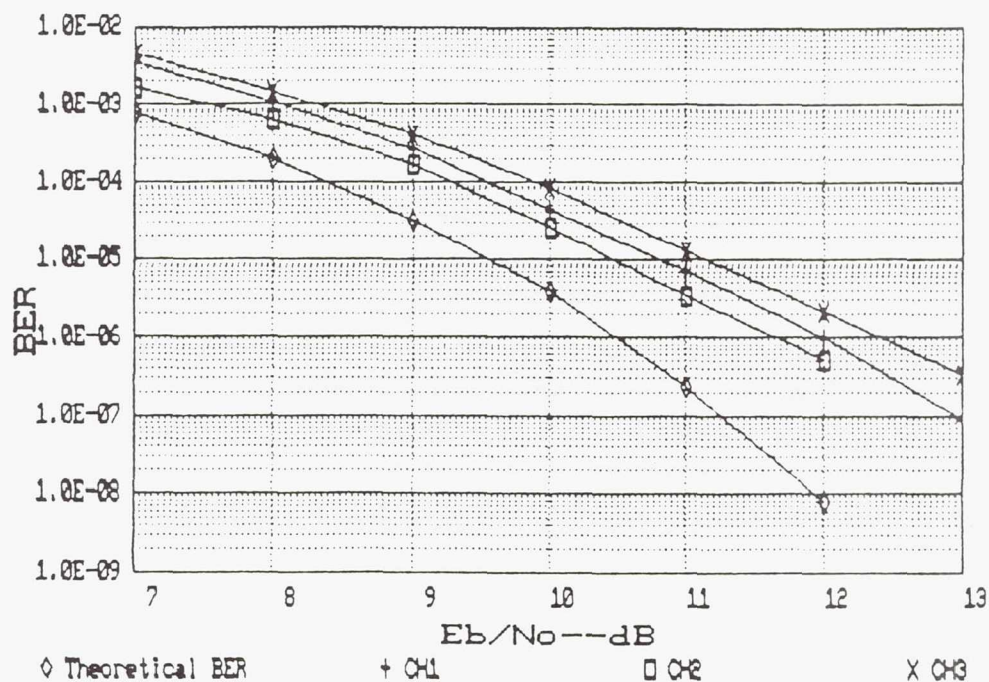◇ Theoretical BER   + CH1   □ CH2   X CH3

FIGURE 7   BER FOR MCDD AT 4.4 MBIT/S
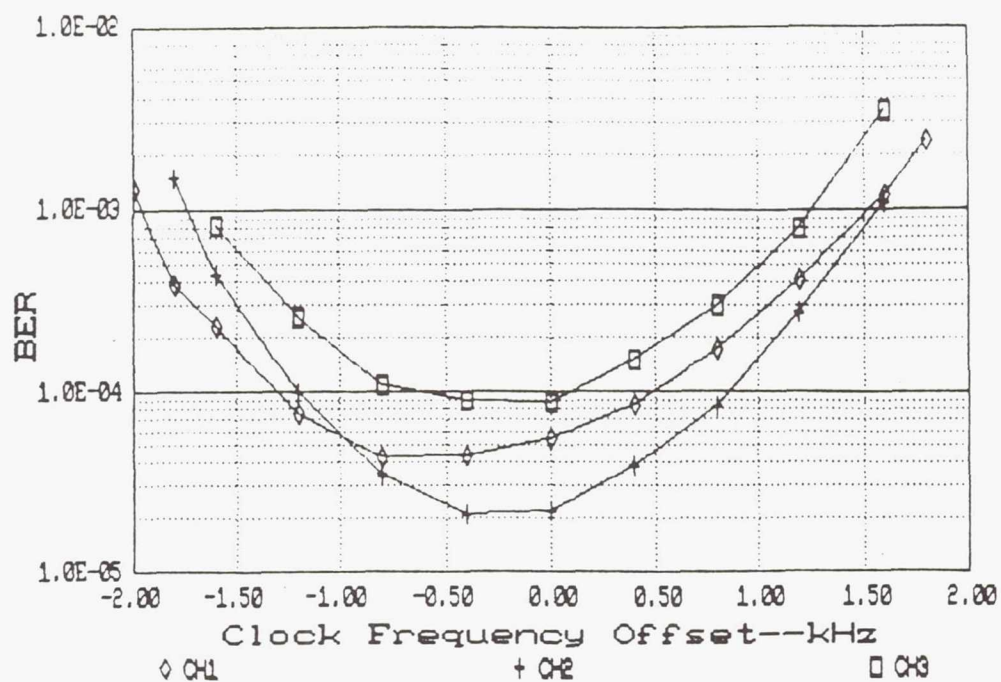
ROW 1, COLUMN 5

121

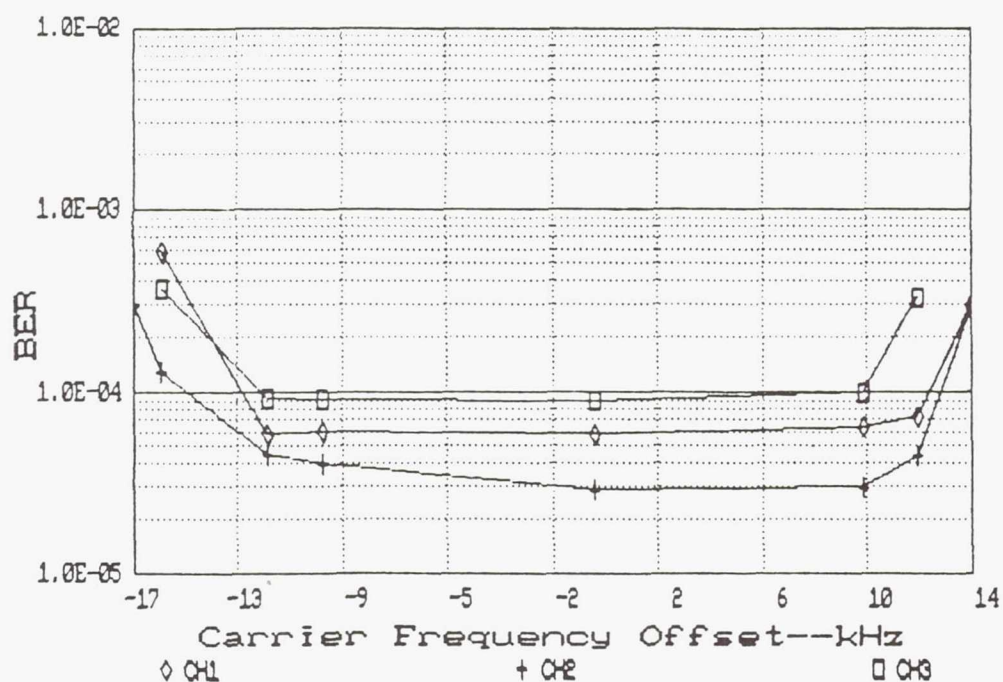FIGURE 8   BER FOR MCDD VERSUS CLOCK FREQUENCY OFFSET

FIGURE 9   BER FOR MCDD VERSUS CARRIER FREQUENCY OFFSET

ROW 2, COLUMN 4

122

TWO ADJACENT CHANNELS (UPPER AND LOWER) ARE THE INTERFERING CHANNELS.
TWO ADDITIONAL ALCATEL SYNTHESIZERS (1 AND 3) ARE USED TO PROVIDE TWO
SOURCE SIGNALS TO THE FDMA SIGNAL GENERATOR FOR THE UPPER AND LOWER
CHANNELS. ANOTHER HP3326A, HP3762A AND DIVIDER ARE USED TO PROVIDE THE
INTERFERING DATA SIGNALS AND CLOCKS TO MODULATE THE UPPER AND LOWER
CHANNEL CARRIERS. MEASUREMENTS ARE ONLY PERFORMED FOR THE CENTER
CHANNEL.

THE DEGRADATION DUE TO THE ACI INTERFERENCE IS NO MORE THAN 0.2 dB LOSS
FOR THE 4.4 MBIT/S DATA RATE CASE.
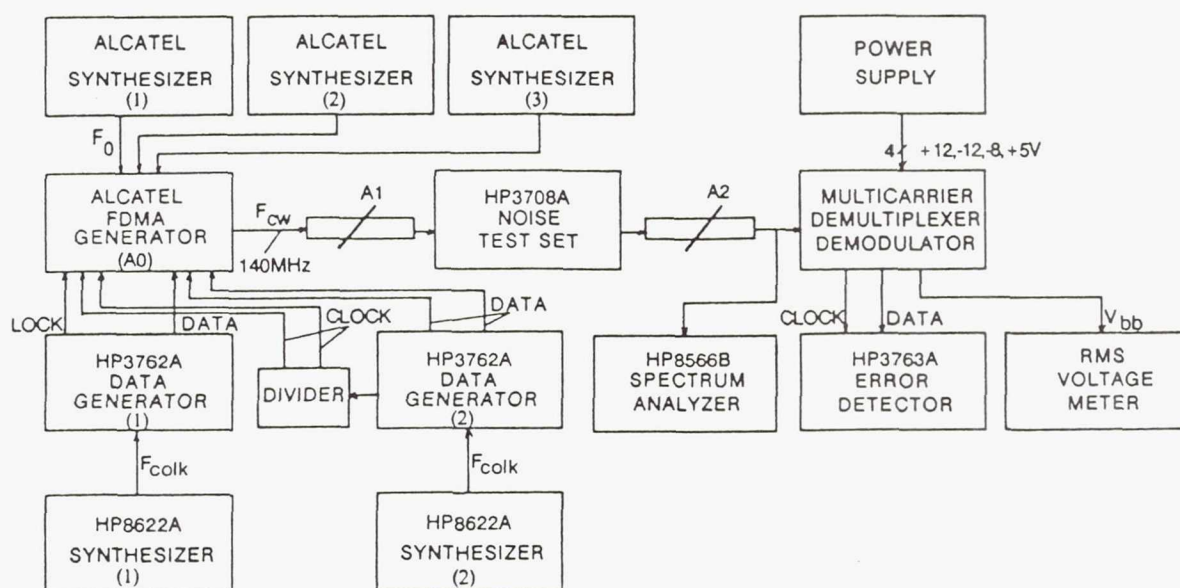
FIGURE 10   SETUP FOR THE ADJACENT CHANNEL INTERFERENCE TEST

ROW 2, COLUMN 5

THE PROOF-OF-CONCEPT HARDWARE FOR THE BBP CONSISTS OF 12 PRINTED WIRED BOARDS: ONE TDM/TDMA CONVERTER, ONE TDMA/TDM CONVERTER, ONE FDMA BUFFER, FIVE FOR THE SWITCH CIRCUITS, FOUR FOR THE CONTROL UNIT.

THE PRINCIPAL FUNCTIONS OF THE BBP CONSIST OF DATA RATE CHANGING; TRAFFIC ROUTING AT BYTE LEVEL, INCLUDING MULTIPLEX (TDM-DOWN), MULTI-CAST AND DISTRIBUTION ETC.; TDM/TDMA CONVERSION; AND DIAGNOSIS.

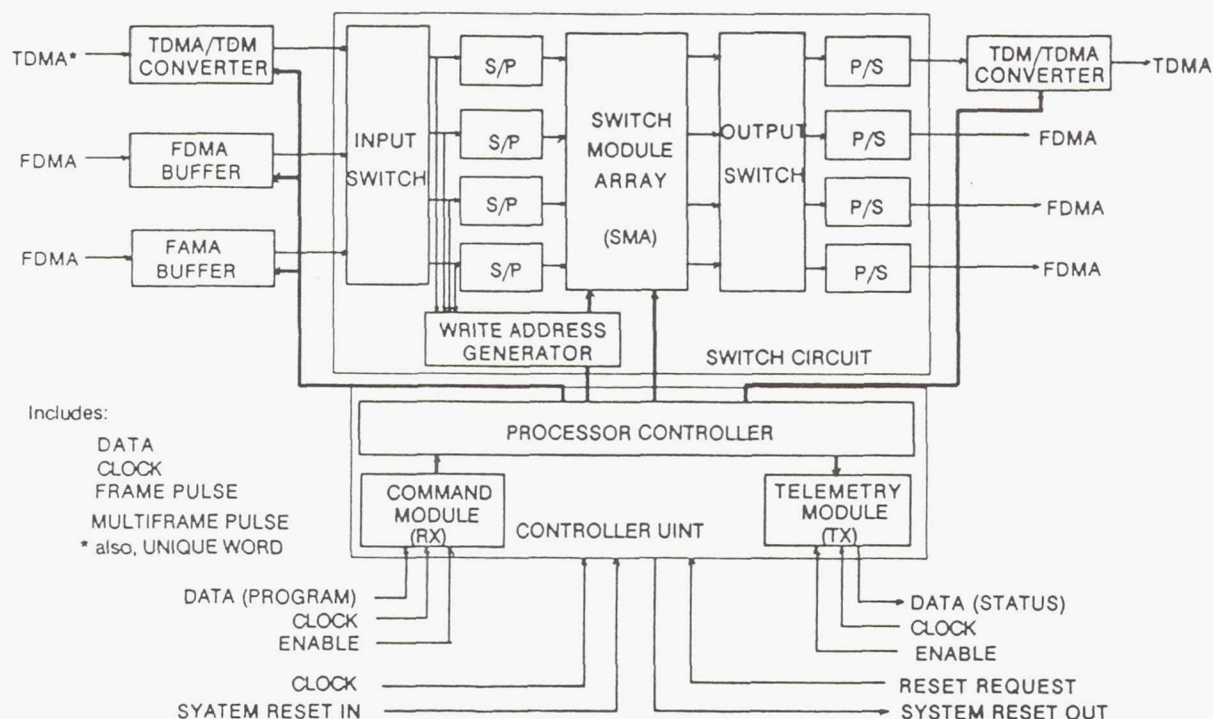THE SWITCH CIRCUIT PERFORMS DATA RATE CHANGES AND ALL SWITCHING FUNCTIONS.

FIGURE 11   BLOCK DIAGRAM FOR BASEBAND PROCESSOR (BBP)

THE BBP HARDWARE IS THE DEVICE UNDER TEST.   THE TEST SET GENERATES THE INPUT SIGNALS FOR THE BBP AND RECEIVES THE OUTPUT SIGNALS FROM THE BBP. THE DATA RATES OF THREE PN DATA ARE 8192, 2048 AND 68.3 KBIT/S RESPECTIVELY.

THE TIMING SIGNAL GENERATOR (A) GENERATES THE MASTER CLOCK OF 16.384 MHZ AND SYSTEM RESET SIGNAL AND OUTPUTS THEM TO THE BBP HARDWARE; (B) GENERATES THREE CLOCKS (8192 KHZ, 2048 KHZ AND 68.3 KHZ) AND SENDS THEM TO THE PN DATA GENERATORS VIA THE MUX; (C) GENERATES SIX KINDS OF FRAME AND MULTI-FRAME PULSES (PERIOD: 250 US, 2 MS, 7.5 MS, 16 MS AND 480 MS) WHOSE PULSE DURATIONS ARE 244 NS; AND (D) OUTPUTS THESE CLOCKS AND PULSES TO THE MULTIPLEXER IN THE TEST SET.

....   ROW 3, COLUMN 3   BBP TESTS
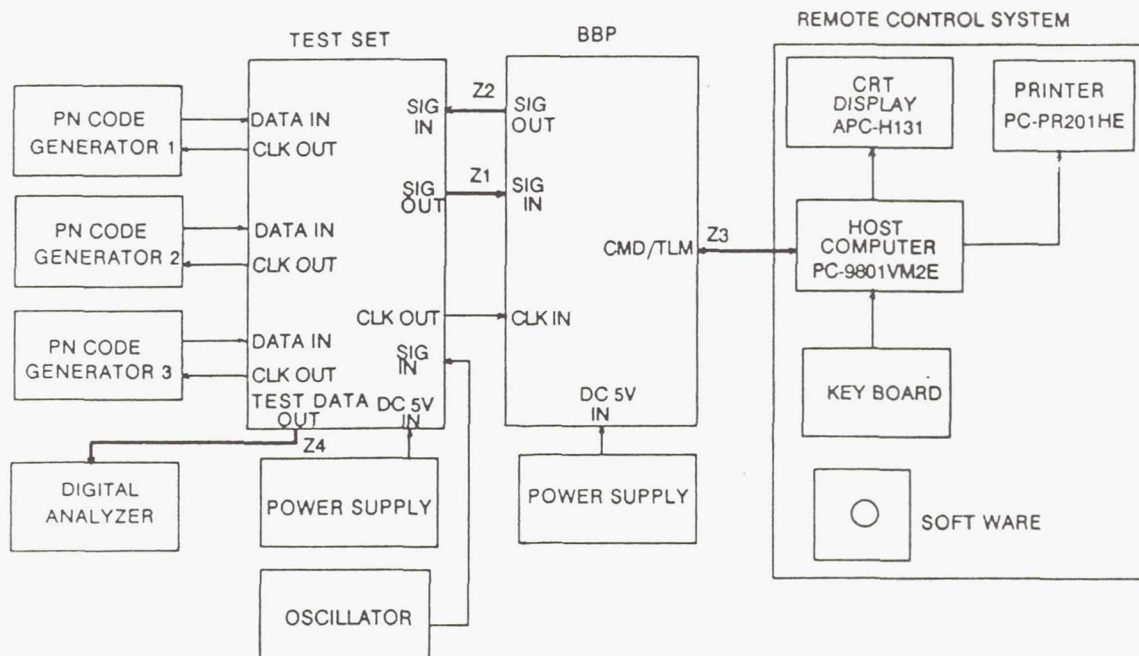
124

FIGURE 12   TEST SETUP FOR THE BBP TESTS

THE MULTIPLEXER GENERATES THREE INPUT DATA STREAMS: TDMA DATA, 1920 KBIT/S TDM DATA AND 64 KBIT/S TDM DATA. THE MUX GENERATES THE UNIQUE WORD SIGNAL FOR EACH INPUT DATA STREAM. THE UNIQUE WORD SIGNALS INCLUDE PREAMBLE DATA AND REFERENCE BURST DATA IN TDMA DATA, MULTI-FRAME DATA INSERTED IN BYTES 0, 16, 32 AND 48 IN TDM DATA.

ANOTHER VERY IMPORTANT FUNCTION OF THE MULTIPLEXER IS INSERTING 8-BIT TEST DATA TO A CHANNEL IN AN INPUT LINE. THE 8-BIT DATA CAN BE ARBITRARILY SELECTED BY SETTING THE "8-BIT TEST DATA" SWITCHES ON THE FRONT PANEL OF THE TEST SET. THE LINE AND THE CHANNEL CAN BE SELECTED BY SETTING THE OUT LINE SEL AND OUT CHANNEL SEL DIGITAL SWITCHES ON THE FRONT PANEL OF THE TEST SET RESPECTIVELY. THE MULTIPLEXER INSERTS THE RECEIVED PN DATA INTO EVERY CHANNEL IN EACH OUTPUT LINES, EXCEPT THE CHANNEL INTO WHICH 8-BIT TEST PATTERN DATA ARE INSERTED.

THE DEMULTIPLEXER RECEIVES THE OUTPUT SIGNALS FROM THE BBP HARDWARE, AND SELECTS ONE FROM 4 OUTPUT LINES AND A CHANNEL IN THE SELECTED LINE TO DISPLAY. THE SELECTION IS PERFORMED BY SETTING THE IN LINE SEL AND IN CHANNEL SEL SWITCHES IN THE FRONT PANEL OF THE TEST SET RESPECTIVELY. THE DEMULTIPLEXER DISPLAYS THE SELECTED 8-BIT DATA BY "IN 8-BIT TEST DATA" LED DISPLAY ON THE FRONT PANEL OF THE TEST SET.

. . . .   ROW 3, COLUMN 4   BBP TESTS

THE OTHER FUNCTION OF THE DEMULTIPLEXER IN THE TEST SET IS PERFORMING THE ERROR DETECTION FOR THE SELECTED 8-BIT DATA. THIS IS PERFORMED BY COMPARING THE 8-BIT OUTPUT DATA DETERMINED BY SETTING THE "8-BIT TEST DATA" SWITCH WITH THE 8-BIT INPUT DATA MONITORED ON THE LED DISPLAY. WHEN AN ERROR IS DETECTED, THE "ERROR" LED DISPLAY ON THE FRONT PANEL OF THE TEST SET TURNS ON AND AN ERROR PULSE IS GENERATED AND OUTPUT FROM THE "ERR PLS OUT" BNC CONNECTOR ON THE REAR PANEL OF THE TEST SET.

THE DEMULTIPLEXER ALSO PERFORMS THE SERIAL / PARALLEL CONVERSION OF THE SELECTED OUTPUT LINE DATA AND OUTPUTS THE PARALLEL DATA FROM THE Z4 (TEST DATA OUT) CONNECTOR ON THE REAR PANEL OF THE TEST SET.

THE REMOTE CONTROL OPERATION SYSTEM SIMULATES THE CONTROL FUNCTIONS OF THE CONTROL EARTH STATION IN A SATELLITE COMMUNICATION NETWORK. ITS MAIN FUNCTION ARE COMMAND SETTING, TELEMETRY DATA DISPLAY, AND TRANSMISSION/ RECEPTION CONTROLS.

IBS/IDR BBP TESTS INCLUDE:
    COMMUNICATION BETWEEN HOST COMPUTER AND THE BBP;
        (COMMAND AND TELEMETRY);
    DATA LOAD-UP AND READ-OUT;
    HARDWARE REDUNDANCY SWITCHING (CONTROL STATUS);
    SWITCHING FUNCTIONS:
        MULTI-CAST, TDM-DOWN, DISTRIBUTION, DATA RATE CHANGE;
    DIAGNOSTIC FUNCTIONS:
        COLUMN CONTROL MEMORY DIAGNOSIS
        AND SWITCHING MODULE MEMORY DIAGNOSIS.

THE IBS/IDR BBP PERFORMANCE MEETS THE FUNCTIONAL SPECIFICATIONS.

THE BASEBAND SWITCH MATRIX IS A 16x16 MATRIX WHICH CONSISTS OF 4 CHIPS REALIZED IN GaAs LSI TECHNOLOGY. THE CHIPS DEVELOPED UNDER CONTRACT INTEL-321 HAVE 16 INPUT CHANNELS AND 4 OUTPUT CHANNELS.

LOW POWER CONSUMPTION (7.78 W) IS ACHIEVED THROUGH THE CONSTRUCTION OF BUFFERED FET LOGIC WITH DEPLETION-TYPE FETS.

THE CHIPS ARE VERY SMALL AND HAVE LOW MASS. THE MASS OF THE DEMONSTRATION UNIT IS 1.25 KG AND IT HAS A SIZE OF 24x18x1.7 cm^3.

EACH CHANNEL RUNS AT A 60-MHz CLOCK RATE; USED IN PAIRS, THEY SUPPORT 120-MBIT/S DATA RATE.

YCP-65968
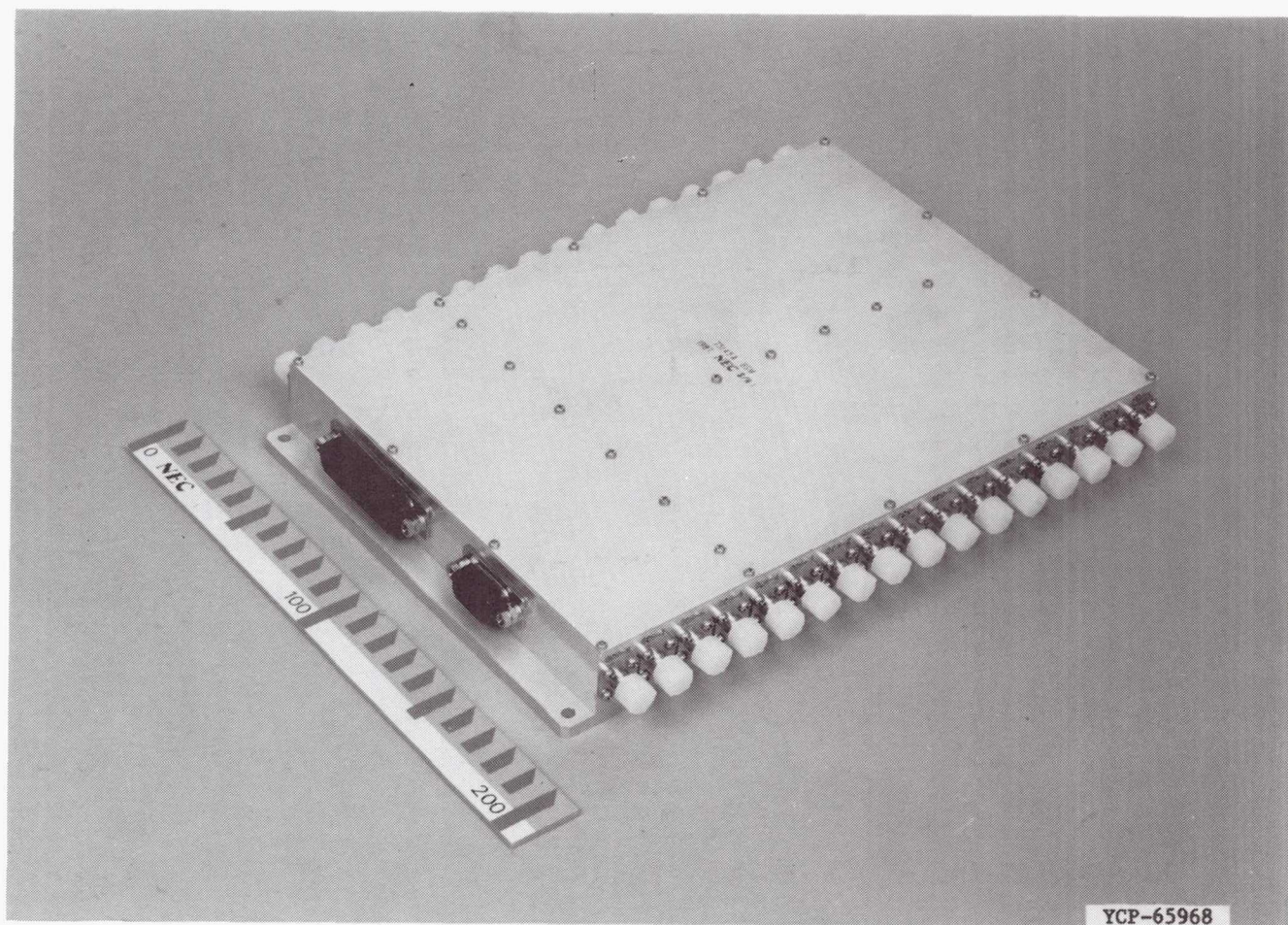
FIGURE 13   PHOTOGRAPH OF BASEBAND SWITCH MATRIX

BSM TESTS   ROW 3, COLUMN 1   ++++

THE BSM CHIP CONSISTS OF 4 FOUR-BIT REGISTER-1s, 4 FOUR-BIT REGISTER-2s, 4-TO-16 DECODERS AND 64 DIGITAL SWITCHES. REGISTER-1 STORES THE SWITCH ADDRESS COMING FROM THE DCU, AND LOADS REGISTER-2 UPON A DCU LATCH PULSE.

REGISTER-2 FEEDS THE 4-TO-16 DECODER WHICH SELECTS ONE SWITCH IN THE 16-GATE COLUMN. EACH SWITCH CONSISTS OF AN AND GATE MADE WITH DEPLETION-TYPE FET. THE AND GATES ARE ARRANGED ON THE CROSS POINTS OF THE MATRIX AND THE SIXTEEN AND-GATE OUTPUTS ARE INTERCONNECTED BY WIRED-OR TO THE OUTPUT LINE.

IN DYNAMIC OPERATION, THE DISTRIBUTION CONTROL UNIT (DCU) CAN UPDATE THE CONFIGURATION OF THE MATRIX UP TO 50 TIMES IN THE 2 MS FRAME. THROUGH THE TT&C INTERFACE, TRAFFIC FLOW PATTERNS ARE STORED IN THE OFF-LINE MEMORY OF THE DCU. AT THE BEGINNING OF THE MASTER FRAME PULSE (8129x2 MS = 16 SEC), THE NEW PATTERN (MAXIMUM 50 CONFIGURATIONS) IS PLACED IN SERVICE IN THE ON-LINE MEMORY OF THE DCU.

BSM TESTS   ROW 4, COLUMN 1   ++++
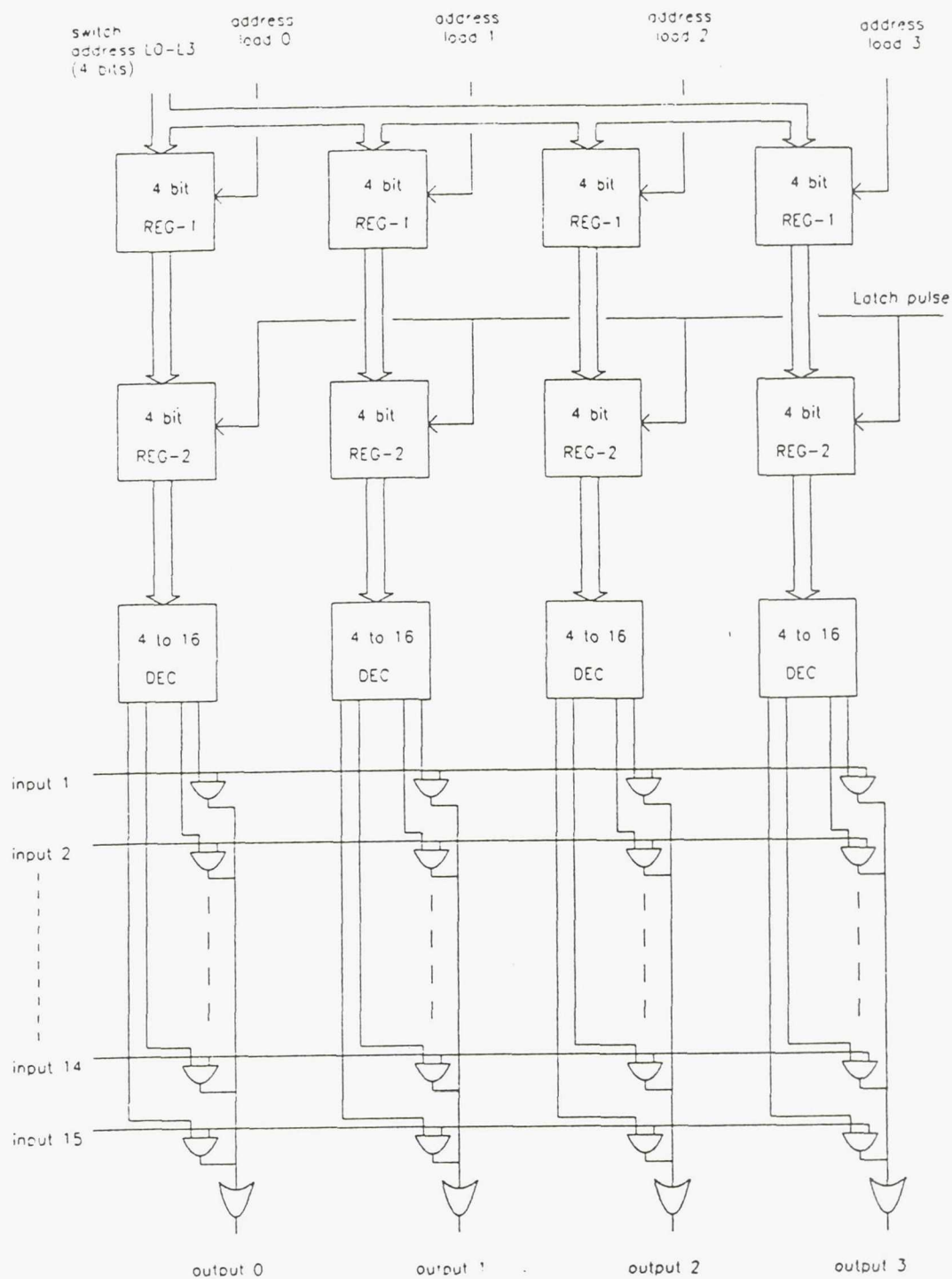
FIGURE 14   GaAs LSI BSM Chip Block Diagram

BSM Tests   ROW 4, COLUMN 2   ++++
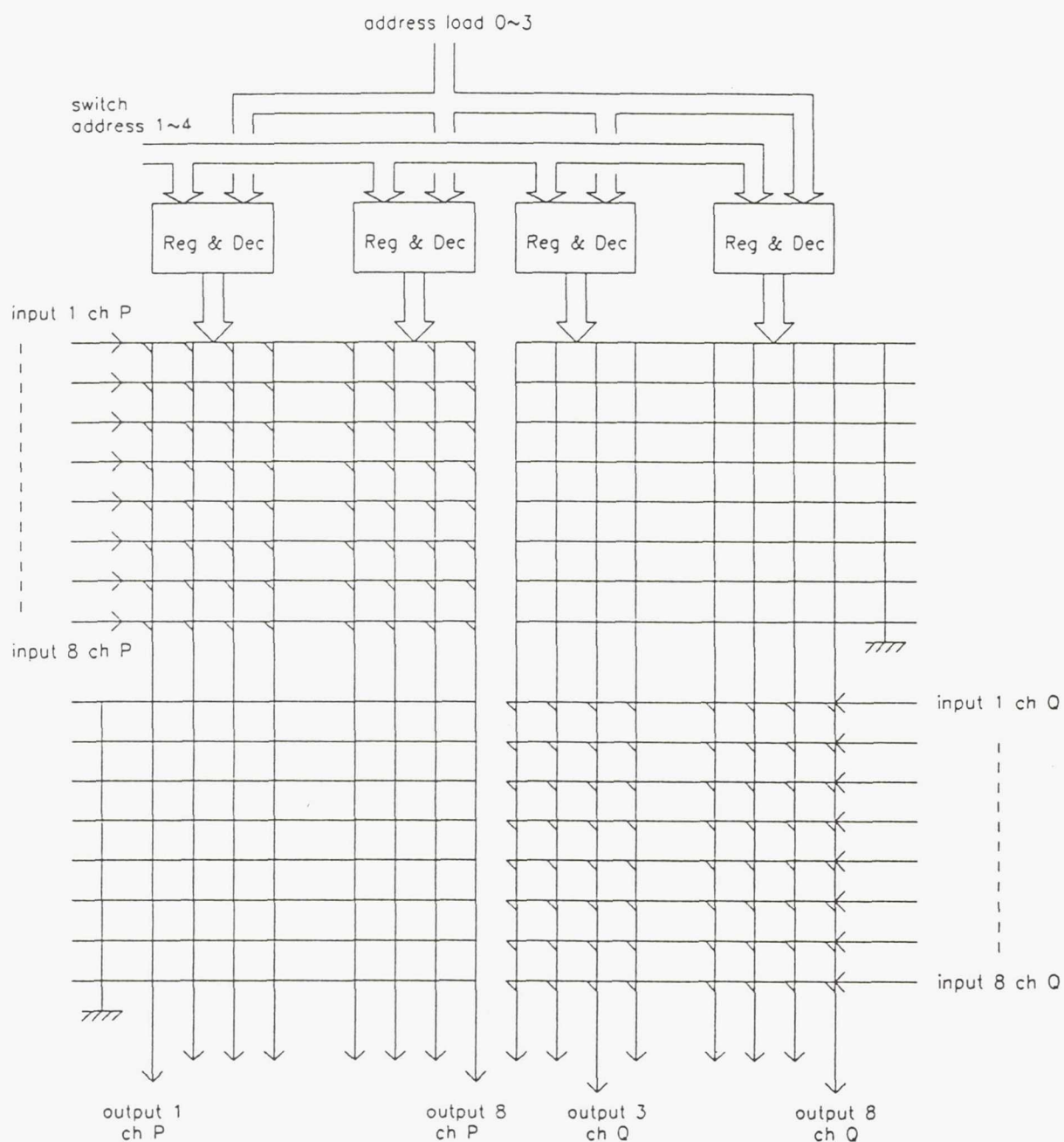
FIGURE 15   BSM TEST IMPLEMENTATION: TWO 8 x 8 MATRICES

IN THE TEST CONFIGURATION, THE BSM SIMULTANEOUSLY CONNECTS DYNAMICALLY
AND STATICALLY TWO SETS OF 8 INPUT PORTS TO 8 OUTPUT PORTS FOR P AND Q
CHANNELS, UNDER THE CONTROL OF THE DISTRIBUTION CONTROL UNIT.
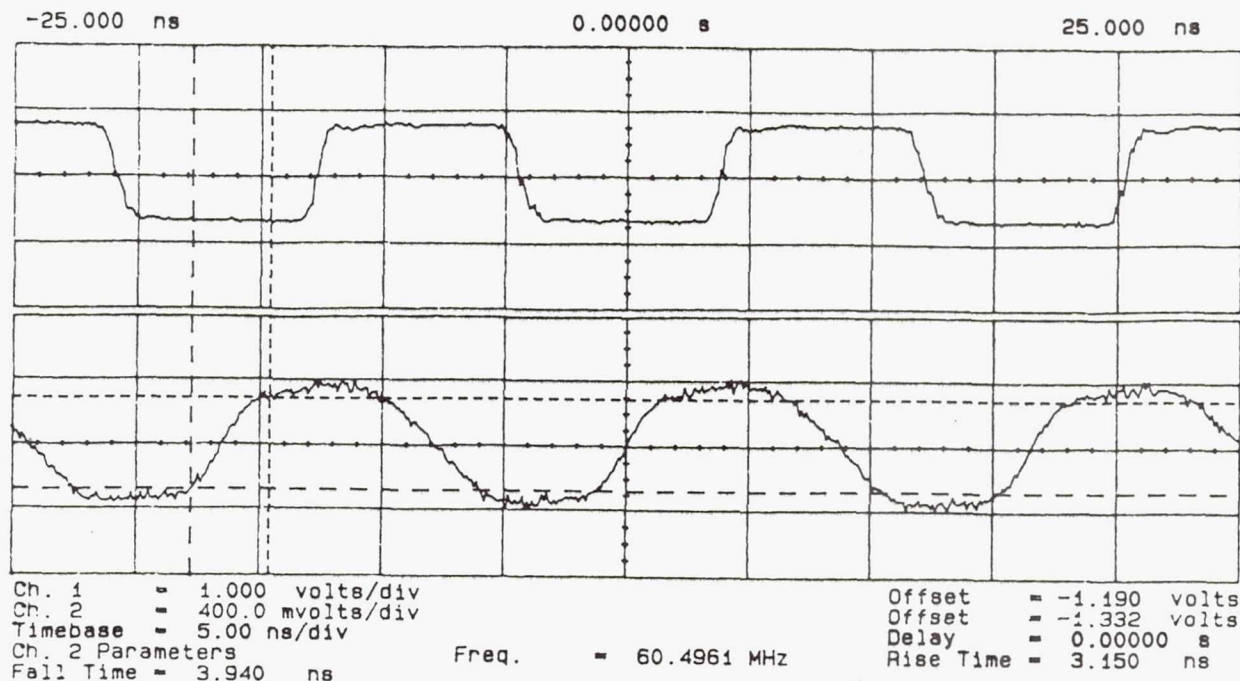[THIS IMPLEMENTATION USES HALF THE SURFACES OF THE CHIPS.]

BSM TESTS   ROW 3, COLUMN 2   ++++

```
 -25.000  ns                    0.00000  s                    25.000  ns
```

```
Ch. 1    =   1.000   volts/div              Offset   = -1.190   volts
Ch. 2    =   400.0   mvolts/div             Offset   = -1.332   volts
Timebase =   5.00 ns/div                    Delay    =  0.00000  s
Ch. 2 Parameters          Freq.   = 60.4961 MHz   Rise Time =  3.150   ns
Fall Time = 3.940   ns
```

FIGURE 16   INPUT AND OUTPUT WAVEFORMS

THE OUTPUT AND INPUT LEVELS ARE ECL COMPATIBLE.
ALL THE SWITCHES FUNCTION AS REQUIRED.

WITH A RANDOM SEQUENCE (LENGTH = 2^{15} - 1) AS INPUT AND AN ERROR
DETECTOR AS A MONITOR, A TEST HAS CONFIRMED THAT THE WORKING SPEED IS
THE 60-MHz CLOCK RATE, AS REQUIRED IN QPSK 120-MBIT/S TDMA SYSTEMS.
SOME SWITCHES WORK UP TO 90 MSYMBOL/S.

WHEN THE BSM OPERATES AT 60 MHz, SOME PRECAUTIONS IN THE RISE TIME AND
THE FALL TIME MEASUREMENTS ARE REQUIRED. TO AVOID LINE REFLECTIONS, WE
ADDED AN ECL GATE WHICH TERMINATES ON A 50-OHM RESISTOR.

BSM TESTS   ROW 4, COLUMN 3   ++++
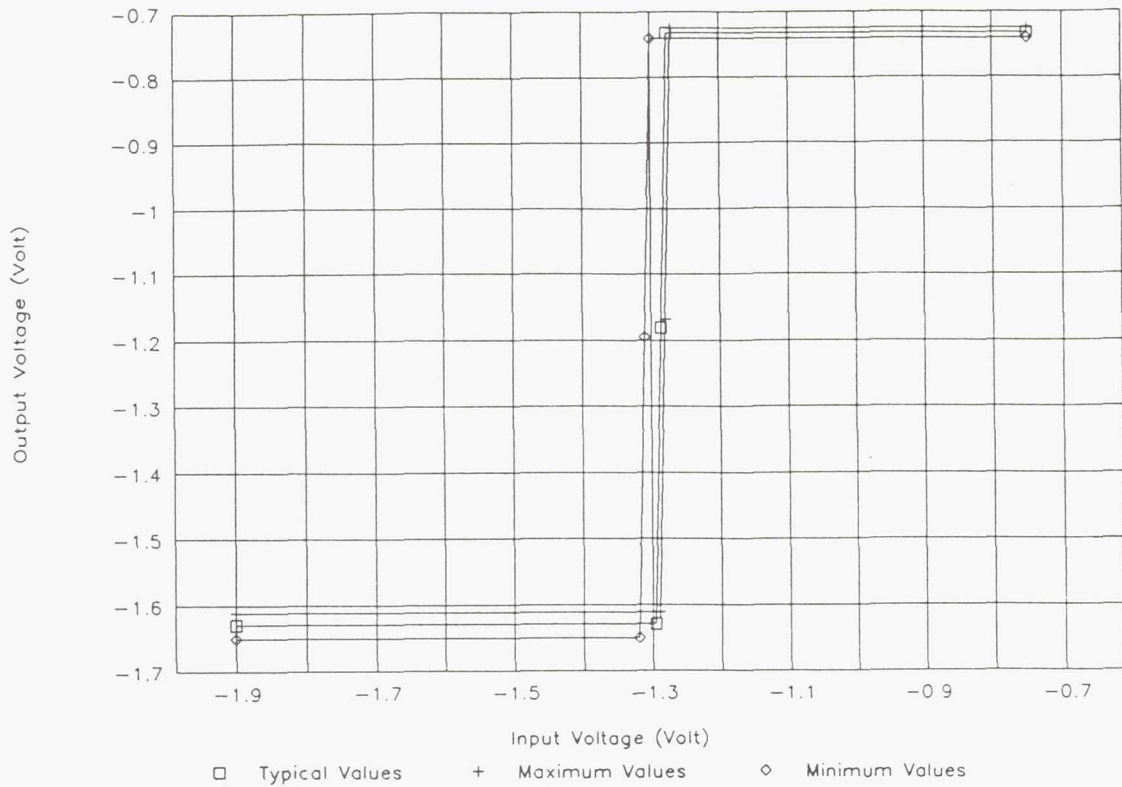
Transfer Characteristics

FIGURE 17   BSM: V_{out} VERSUS V_{in}

THE FIGURE SHOWS A COMPOSITE OF OBSERVED INPUT / OUTPUT CHARACTERISTICS
FOR MANY BSM GATES.

THE INPUT / OUTPUT TRANSFER CHARACTERISTICS
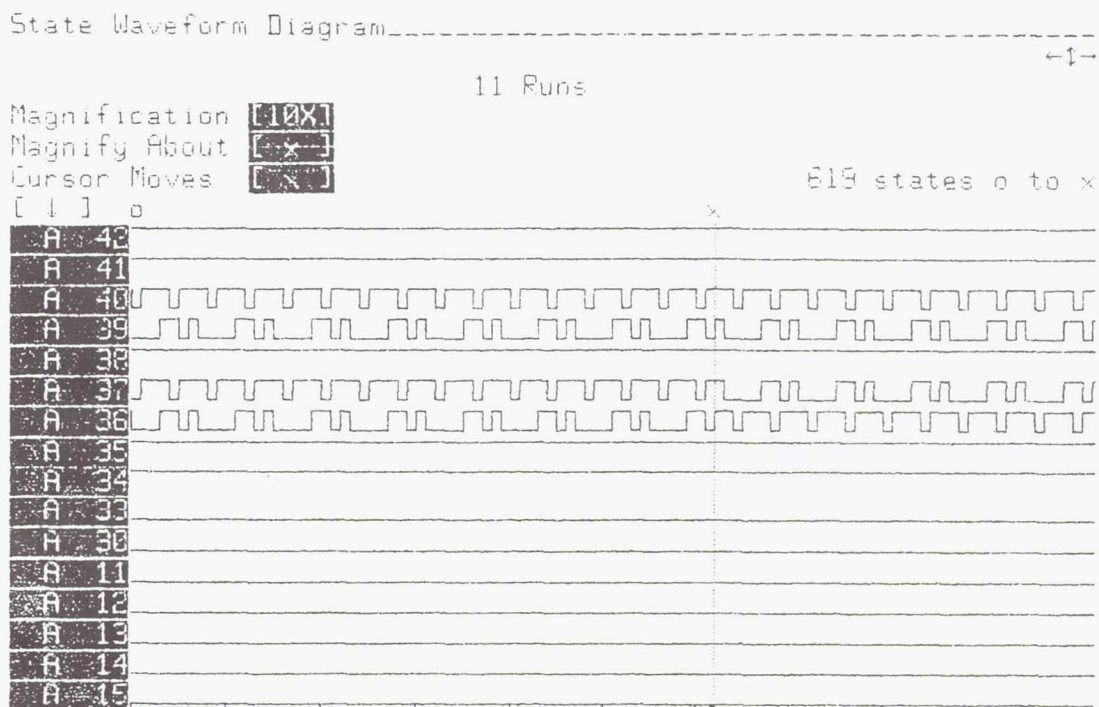FOR ALL GATES ARE ACCEPTABLE.

BSM TESTS   ROW 4, COLUMN 4   ++++

```
State Waveform Diagram_____
                                                              ←1→
                            11 Runs
Magnification [10X]
Magnify About  [-x-]
Cursor Moves   [ X ]                        619 states o to x
[ ↓ ]  o                              x
```


FIGURE 18   DYNAMIC TEST OF THE BSM

THE FIGURE ILLUSTRATES THAT THE DATA STREAMS ARE SWAPPED
FROM P CHANNEL 1 TO P CHANNEL 8.

THIS DYNAMIC TEST SHOWED THAT THERE ARE NO BIT LOSSES
DURING THE SWITCHING OPERATION.

REFERENCE

1.   P.P. NUSPL AND G. DONG, "ON-BOARD PROCESSING FOR
      TELECOMMUNICATIONS SATELLITES", TO BE PRESENTED AT SECOND NASA
      SPACE COMMUNICATIONS TECHNOLOGY CONFERENCE, ONBOARD SWITCHING AND
      PROCESSING, CLEVELAND, NOVEMBER 1991.

BSM TESTS   ROW 4, COLUMN 5   ++++

Page intentionally left blank

GETTING EXPERT SYSTEMS OFF THE GROUND:
Lessons Learned from Integrating Model-based Diagnostics with Prototype Flight Hardware

Amy Stephan
Carol A. Erikson

TRW Space & Technology Group
One Space Park
Redondo Beach, CA 90278

ABSTRACT

As an initial attempt to introduce expert system technology into an onboard environment, a model-based diagnostic system developed using the TRW MARPLE software tool was integrated with prototype flight hardware and its corresponding control software. Because this experiment was designed primarily to test the effectiveness of the model-based reasoning technique used, the expert system ran on a separate hardware platform, and interactions between the control software and the model-based diagnostics were limited. While this project met its objective of demonstrating that model-based reasoning can effectively isolate failures in flight hardware, it also identified the need for an integrated development path for expert system and control software for onboard applications. In developing expert systems that are ready for flight, we must evaluate artificial intelligence techniques to determine whether they offer a real advantage onboard, identify which diagnostic functions should be performed by the expert systems and which are better left to the procedural software, and work closely with both the hardware and the software developers from the beginning of a project to produce a well-designed and thoroughly integrated application.

INTRODUCTION

This paper discusses research at TRW aimed at integrating artificial intelligence (AI) technology into an on-board environment. The work described was a joint effort of the Expert Systems on Spacecraft internal research and development project, spacecraft power system engineers, and flight software developers. The goal of the project was to demonstrate an expert system working in concert with on-board flight software and a hardware testbed. This paper first discusses the nature of AI and flight software, and issues involved in integrating these two technologies. We then describe the work performed at TRW and the results of the expert system tests on the prototype power subsystem. We conclude with a discussion of the problems encountered and lessons learned from this work, and describe a methodology for future integration of AI and onboard data systems.

FLIGHT SOFTWARE AND EXPERT SYSTEMS

The design and development of onboard flight software is driven by the limited capabilities of onboard hardware, the high reliability required to ensure successful operations in space for extended periods, and the deterministic response times inherent in spacecraft control algorithms [Filarey 90]. Required to operate in an extreme environment, onboard data processing systems are constructed of specialized parts designed to survive radiation effects while minimizing size, weight and power consumption. These systems do not provide the throughput and memory capabilities found in ground-based computer systems.

The software developed for onboard systems must also maintain a very high level of reliability, and this places a number of constraints on flight software design and development. Flight software must be designed to be testable; its reliability must be proven on the ground well before launch. The tests performed on flight software range from low-level unit testing, in which every machine instruction and every path of each software module is executed and tested, to spacecraft integration and testing, where the entire spacecraft is assembled and tested. To ensure that the software can be properly tested, its design must be deterministic: given a set of inputs, one must be able to pre-determine both the required output and the path the software will take to produce that output. Most spacecraft control algorithms involve sampling spacecraft sensor data, analyzing the data and sending commands to spacecraft units to maintain stabilized control. The performance of these algorithms depends on accurate scheduling based on known delays between sensor samples and commanded responses. Given these constraints, onboard flight software is limited to those spacecraft functions which are deterministic, efficient in memory use, executable within a guaranteed time interval, and testable.

Typical candidates for onboard flight software include basic spacecraft support functions such as attitude control, thermal control and power management, and spacecraft fault detection and management. Onboard fault detection is limited to a set of predetermined faults which can be diagnosed by a simple analysis of available spacecraft sensor data. While in some cases the flight software may be able to isolate the source of a fault and switch to a redundant unit or alternative control scheme, often the flight software reacts to an anomaly by placing the spacecraft in a non-operational "safe-hold" mode, relying on the ground to isolate and correct the fault. Ground operators also control all mission planning and operations, often through detailed, low-level command sequences. As spacecraft missions require greater

survivability, autonomy and complexity, this heavy reliance on ground support must be alleviated by software that can perform higher-level decision making [Fesq 89].

Research in the field of AI has sought to increase the capabilities of on-board software in the areas of diagnosis, planning and scheduling. The potential benefits of such research are many. Onboard diagnostic and planning would allow spacecraft to achieve a high level of autonomy, operating for months without ground contact. More complex in-flight navigation could be achieved with little ground control. Large systems such as the Space Station Freedom could make better use of available resources. In addition to increasing the satellite's on-orbit capabilities, enhanced onboard software could significantly reduce the cost of ground operations.

While AI has made much progress in the years since rule-based expert systems first became popular, these advances have not generated a great deal of interest among flight software developers. This results partially from the fact that AI researchers are seldom the same people who develop flight software, and the two groups have markedly different approaches to software development. AI systems are almost always built on specialized workstations where memory and throughput limits seldom impact performance. Even more importantly, AI systems typically are designed to exhibit novel behavior and respond to uncertain conditions. Rule-based systems, for example, are designed to be able to chain through rules in unexpected ways, and often several paths may exist between a set of input and output data. By their nature such systems are non-deterministic and difficult to test. It is no wonder, then, that when AI researchers emerge from their labs with their latest prototype, the onboard software community issues a collective yawn.

The fields of AI and flight software are both changing, however, and the goal of our research is to understand how these two software methodologies can be combined to more effectively carry out the spacecraft mission. More deterministic expert system techniques, such as model-based reasoning, are now in use operationally. A growing sub-field of AI is actively researching methods for developing testable applications. In areas outside of onboard processing, including commercial applications and ground software, AI has been successfully integrated with conventional software. In fact, in the ground-based domain, AI is fast becoming just another tool in a programmer's repertoire. As more powerful processors and larger memories migrate onboard, the flight software environment will no longer be so highly constrained and onboard software will be able to take advantage of applicable AI techniques.

SYSTEM DEVELOPMENT

As an initial attempt to introduce expert system technology into an onboard environment, a model-based diagnostic system was integrated with prototype flight hardware and software. This work demonstrated the ability of a model-based expert system to isolate hardware faults. It further showed that an expert system could be effectively integrated with conventional flight software to produce a significant improvement in diagnostic capabilities. For purposes of this demonstration, the testbed hardware and software were also required to operate independently of the expert system. We therefore chose a loosely-coupled software architecture, which limited the amount of communication and cooperation possible between the flight code and the expert system. While this system met all of our initial goals, its limitations have taught us that tighter integration between conventional and AI-based code is needed for a realistic onboard system.

The prototype flight hardware and software illustrated in Figure 1 were developed as part of an advanced concept power subsystem project. The hardware testbed included power control electronics, a solar array simulator, spacecraft batteries, current and voltage sensors, and a 1750A instruction set architecture (ISA) onboard processor. While the testbed was being assembled, the associated power control software was developed in 1750A assembly language on an HP9000 workstation. The power control software included sensor processing functions, a control algorithm to ensure constant battery charge while preventing overcharge, command output functions, and limited fault management capabilities. Once the testbed was assembled and the flight software was completely developed, the code was downloaded to the onboard processor and the hardware and software were integrated and tested.

Although the expert system was planned as part of the demonstration system from its inception, the flight software and the expert system were considered two separate efforts, each with its own development team. The expert system was a model-based diagnostic system developed on a Texas Instruments microExplorer using MARPLE, an in-house model-based expert system shell [Cowles 90; Fesq 91]. The MARPLE shell implements a model-based reasoning technique known as constraint suspension. This technique, developed in the MIT AI Lab, does not attempt to model the behavior of failed components. Instead, it constrains the values placed at system components based on a series of transfer functions and systematically suspends these assumptions to isolate a failed component [Davis 88].

MARPLE is a LISP-based tool containing all the functions necessary to run a model-based diagnostic system: the user need only supply models of the system to be monitored. For this experiment, we developed hierarchical models of the power testbed, including a top-level model of the entire testbed and more detailed models of the power distribution electronics and the solar array simulators. Graphical representations of these models were developed for the expert system user interface shown in Figure 2. After the models were coded using MARPLE's high-level model definition language, the expert system was tested using a VAX-based power system simulator. This simulator allowed us to test the operation of the expert system before integration with the hardware testbed, and also allowed us to exercise the system against a wider range of faults than was possible on the actual hardware.

Communication between the flight software and the expert system was designed to be minimal and to be compatible with the flight software's existing command and telemetry capabilities. A one-way communication scheme was designed in which the flight software periodically sent packets of information containing spacecraft sensor data to the expert system. The expert system converted this raw data stream into voltage and current readings before processing it through the MARPLE models. A full-duplex mode of operation was planned but never executed, in which the expert system would send messages back to the flight software, indicating which unit had failed and allowing the flight software to bypass the fault. The diagnostic capabilities of the expert system were not designed to be integrated with the flight software, but were intended to augment or replace the flight software's fault management capabilities.

values and data were collected. This information was analyzed and the expert system models were modified appropriately. Most modifications involved changing scale factors in the model's equations, although in the case of the power control electronics, a new set of models had to be developed to reflect the component's actual behavior. Data analysis and off-line calibration lasted about two months. This period could have been shortened if contention for use of the testbed had not delayed the data collection process.

After the off-line calibration was complete, we linked the expert system to the flight software and began on-line calibration. An RS-232 link was established between the spacecraft processor and the microExplorer. Every 25 ms the power control software sent testbed sensor data to the expert system. The model-based system was first tested with nominal testbed data (no faults) to assure that it was correctly monitoring the testbed performance.
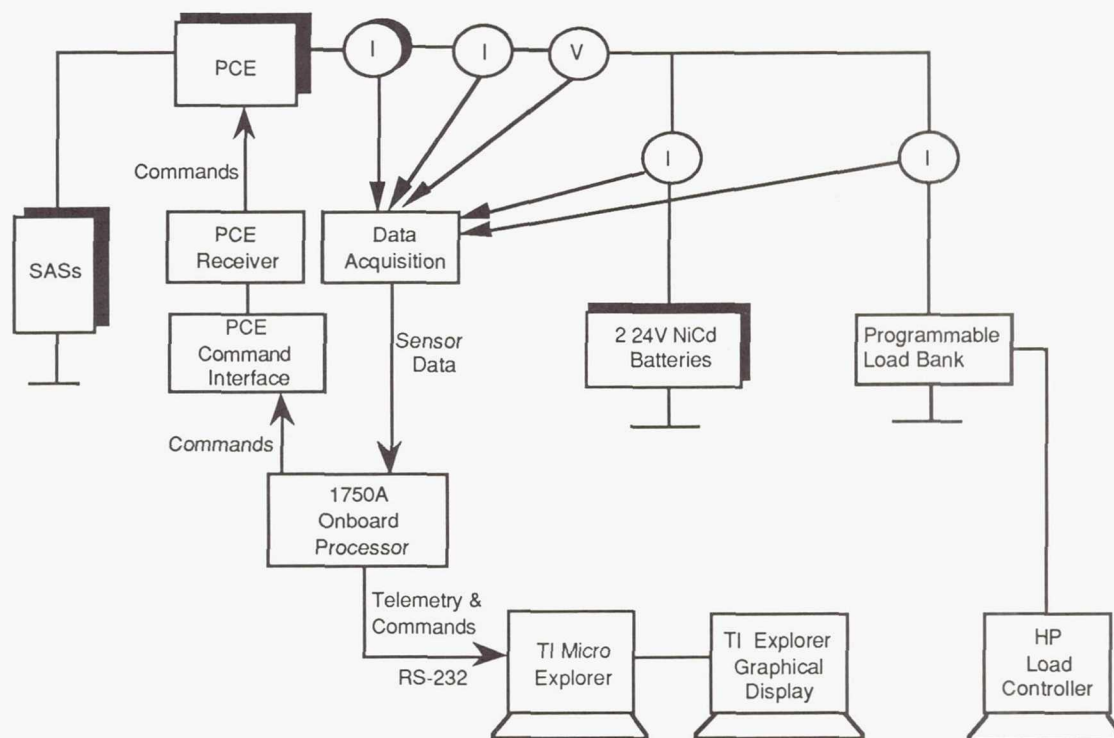
Figure 1. The Advanced Concept Power Subsystem Testbed

Only after the flight hardware, software and expert system were fully developed and tested did we begin integrating the expert system with the prototype power subsystem. The first step in this integration was the calibration of the expert system models to the hardware testbed. Because the simulator used to perform the initial tests on the expert system was developed before the test hardware was available, it was based on idealized models of the power system behavior. Needless to say, these theoretical simulations did not always adequately reflect the actual performance of testbed components. Once the testbed was operational, each component was exercised through the range of its

Although the expert system models were not extremely accurate (most components were modeled within one to three percent of their actual values), the expert system was able to monitor the testbed's performance without producing any false alarms. The expert system was now ready to diagnose failures.

Prior to integration, we identified several typical on-orbit faults which could be safely simulated on the hardware. The faults tested included open and short circuit failures in the power control electronics, open circuit battery cells, shadowed solar arrays and failed sensors. These faults were injected into the testbed during a normal

137

system run and the expert system's performance was monitored. In most cases, the expert system was able to diagnose the fault immediately, displaying its conclusions on the graphical user interface.

RESULTS

Through a series of extensive tests, we proved that the model-based technique is capable of diagnosing hardware faults. Table 1 summarizes the faults that we injected in the testbed and the ability of the expert system to isolate the source of each failure. A majority of the test cases involved open or short circuit failures in the power control electronics since these failures may be easily injected in the hardware by removing a fuse, or by connecting the output of a solar array simulator directly to the output of the power control electronics. Over the range of operation for which the expert system was calibrated, it diagnosed the correct unit in two-thirds of the open circuit test cases, and isolated the fault to correct string of the power control electronics in the remaining cases. Additional open circuit tests were performed in the battery overcharge region, for which no calibration data was available. As might be expected, the diagnostic system fared poorly in this region, correctly identifying less than 50 percent of the faults. In the short circuit tests, however, the expert system performed well in both the overcharge and normal operation modes. In both of these regions of operation, the model-based technique was able to correctly identify the faulty string each time a short circuit failure was induced. In 66 percent of the short circuit test cases, the expert system was able to isolate the fault to the component level.

Tests were performed with failures injected in the solar array simulators to verify the ability of the expert system to distinguish between solar array and power control electronics failures. Open, short and shadowed solar array cells were induced in the testbed by programming the solar array simulators with the skewed characteristics of these failures. Since simulation was the the only means of causing these failures, test cases were not as extensive as with hardware induced failures. The expert system was able to isolate a failed string of the solar array for each type of fault induced. Although successful, test cases of battery failures were necessarily limited because of the inherent danger of introducing short or open circuits in battery cells. (An explosion of a 48V battery would have been detected by the expert system, but the test conductors may not have survived to document the results.)

Although the flight software and the expert system development efforts were segregated, communication
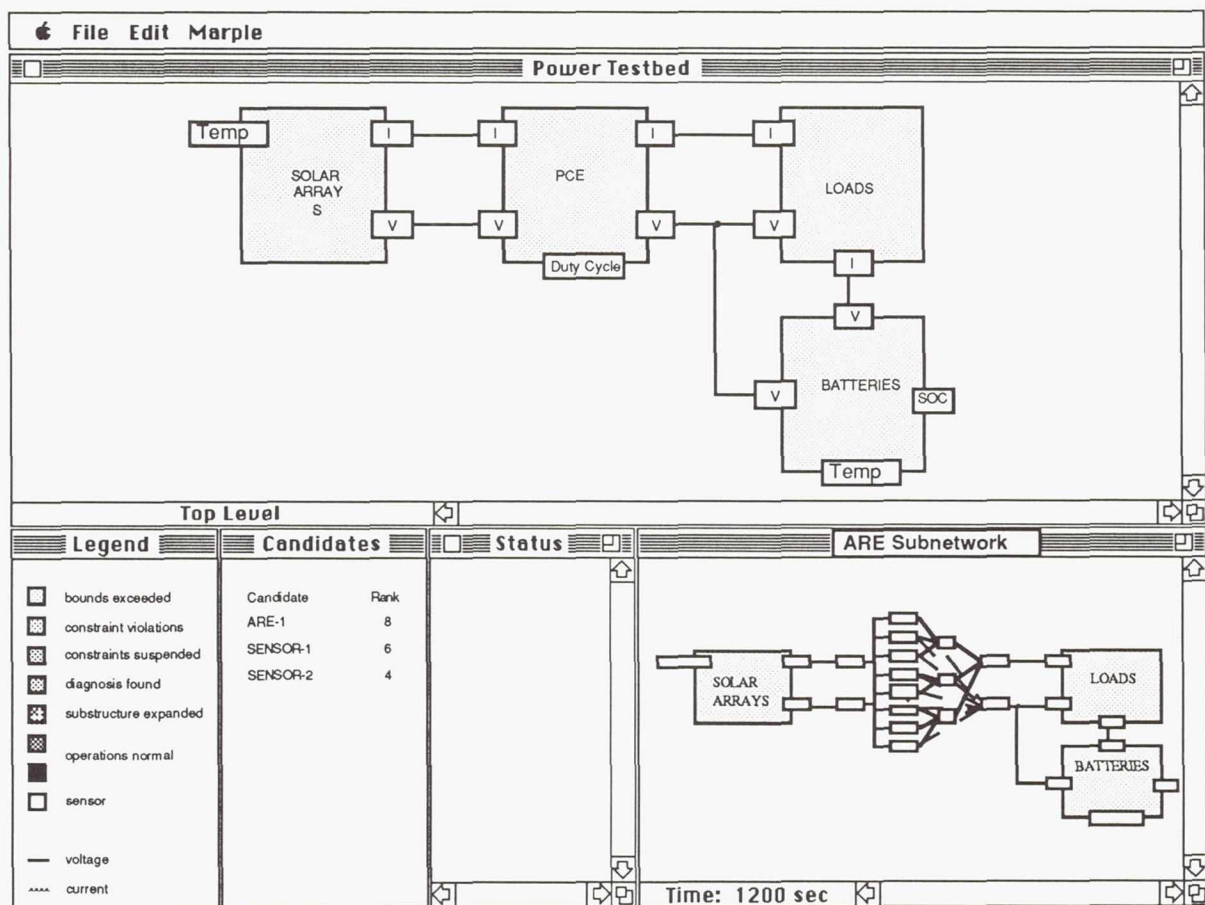

Figure 2. MARPLE User Interface developed for the advanced power testbed.

between the hardware, software and expert system personnel was strong throughout the life of the project. The expert system developers participated in weekly project meetings and were co-located in the same lab as the hardware and flight software developers. The testbed engineers, although busy, were interested in the expert system project and willing to answer questions. This close communication contributed to the relative ease with which the expert system was integrated with the hardware and flight software.

Another factor which helped smooth the integration effort was the extensive off-line calibration performed on the expert system using testbed data. Long hours were spent acquiring performance data, analyzing the results, and adjusting the expert system models as necessary. Having models which accurately represented the hardware proved invaluable in reducing the schedule required to test the expert system. From the start of on-line testing, the expert system was able to model the behavior of the testbed, allowing us to proceed with the test cases rather than refining the models for each fault injected. The importance of off-line calibration is apparent in our test results. The expert system was much less accurate in regions for which inadequate calibration data was available, such as at very high battery voltages.

The competition for time on the testbed hardware made off-line calibration even more critical. The hardware designers needed the testbed to validate their design and acquire performance data; flight software developers were required to verify the execution of their control algorithms on the actual hardware. Expert system developers were allocated any remaining time for calibration and testing. Conducting the data analysis and model adjustment off-line limited system contention and allowed the hardware and software developers to complete their tasks with little impact from the introduction of the expert system.

The fault diagnosis capabilities of the expert system were a significant improvement over those designed for the flight software. The flight software fault management capabilities were limited to reacting to faults by switching on additional circuitry: while it could identify broad classes of faults and initiate recovery strategies, the flight software could not identify exactly which component had failed. The expert system was able to identify not only what component failed, but in some cases it could characterize the nature of that failure (for example, a short-circuit versus an open-circuit failure in the power control electronics). The expert system also proved more adept at diagnosing sensor failures, immediately identifying the failed sensor, as opposed to the lengthy (and inconclusive) isolation algorithm implemented in the control software.

Although the expert system demonstrated more advanced fault detection and isolation capabilities than the flight software, much of the fault management processing performed by the two systems proved to be redundant. For demonstration and testing, this redundancy served a useful purpose; it allowed us to compare the performance of the expert system to the standard fault detection and isolation techniques used by the flight software. While some redundancy might be beneficial in an onboard system, the overlap between our expert system and flight software reflected the segregated development efforts and was too inefficient for use onboard. If the design of the two systems had been more coordinated, the fault diagnostic capabilities of the expert system could have become an integral component of the flight software's fault management scheme. Instead, the flight software developers designed a standard technique of analyzing spacecraft sensor data and determining faults while the expert system developers analyzed the same data using model-based reasoning with the same end goal -- diagnosing failures in the hardware.

By the completion of our research, we had made significant progress in proving the ability of model-based reasoning to diagnose hardware failures. This success was based on the development of two distinct working systems -- a prototype flight hardware and software system which successfully demonstrated a new concept in power control, and a model-based expert system which could isolate faults in the power system. Any attempt to integrate both systems on a single spacecraft processor, however, would require significant redesign of both systems. The control software would have to be modified to take full advantage of the expert system's fault diagnostics and ensure that data required by both the expert system and flight software were properly managed. The expert system would have to be streamlined to run on a target machine with limited throughput and memory and would have to be converted to a language supported by the spacecraft processor.

DISCUSSION

Although we tested our expert system extensively against both the software simulator and the actual testbed, these tests did not begin to meet the rigorous standards required for flight software. Because it was designed as a demonstration, not a flight system, low-level unit verification tests were not performed on the expert system. Given the relatively structured nature of the model-based reasoning approach used, unit testing could have been successfully performed on most of the expert system code. It is doubtful, however, that all possible execution paths could have been tested under realistic project budget constraints. Similarly, while the expert system performed well against both the simulator and the hardware, its reliability was not proven to be appropriate for onboard use. Again, it is important to realize that this system was designed as a demonstration. An onboard application would have been designed to achieve a higher level of reliability, and calibration and testing would have been more extensive. However, while more thorough testing and calibration undoubtedly would have increased the system's reliability, the size and potential number of execution paths through the expert system would have prevented standard testing methods from adequately measuring this reliability.

Current research at TRW seeks to develop new verification and reliability assessment methods for this type of model-based system based on a mathematical analysis of the expert system's properties. Additionally, the expert system used for this project has been redesigned and recoded in Ada, and hosted on a 1750A ISA processor. While such work eventually may prove fruitful for introducing large scale flight-based expert systems, our research suggests that such large, loosely coupled applications may not be the most effective use of AI onboard. For many applications, a tighter coupling of AI and procedural code tailored to the particular problem at hand would be more effective. In such a system, traditional procedural code would be enhanced by small, embedded applications of AI-based techniques. Because these small model- or rule-based modules would have a limited number of execution paths, they could be verified through standard methods. Such an approach could be used to avoid unnecessary redundancy and capitalize on the strengths of each programming technique. This embedded approach would also eliminate the need to redesign and rehost a ground-based prototype to run on a flight processor. If AI sections of code were designed and developed as part of an

integrated flight software development effort, such problems as processor contention, communication, and integration could be simplified and a more powerful flight software package would result.

At the conclusion of our project, a team of expert system and flight software developers revisited the control software requirements and design. As part of this effort, the AI personnel conducted a detailed study of the flight software's goals and design, and the flight software engineers assessed the strengths and limitations of the expert system. The team then worked together to identify several sets of requirements which might be better served by AI-derived techniques than by procedural code. These included battery recharge ratio adjustment, optimizing peak power tracking, sensor consistency checking, performance prediction and monitoring the pattern of controller faults. In each of these cases, the AI-technique would have been more efficient, more accurate or more flexible than the procedural implementation. Also, each of these mini-applications of AI would be well-defined, so that its performance could be thoroughly characterized through test procedures. Were this sort of analysis to be

Fault Induced	Number of Trials	Test Method	Expert System Diagnosis	Remarks
PCE FAULTS				
Open Circuit	8	H/W	6 PCE, 2 BATT	Two BATT from overcharge
Short Circuit	6	H/W	PCE	
Command I/F	1	H/W	PCE	
BATTERY FAULTS				
Low Discharge	10	VAX SIM	BATT	H/W Test may damage battery
Open Cell	1	H/W	BATT	
Short Cell	1	H/W	BATT	
Overtemp	10	VAX SIM	BATT	No temp sensor on testbed
S/A FAULTS				
Open Cell	1	H/W SIM	S/A	
Short Cell	1	H/W SIM	S/A	
Shadow	10	H/W SIM	S/A	
Temp Increase	10	VAX SIM	S/A	
SENSORS				
Stuck on	1	H/W	SENSOR	Current/voltage sensors
Stuck off	1	H/W	SENSOR	
LOAD FAULTS				
Disable/Enable	10	VAX SIM	LOAD	No i/f with loads on H/W
Change sizing	10	VAX SIM	LOAD	
S/A THREATS				
Pellets	10	VAX SIM	S/A	No temp sensor on testbed
Laser Attack	10	VAX SIM	S/A	

Note: All test cases performed on the hardware testbed were also successfully executed with the VAX-based simulator.

Table 1. A Summary of Model-based System Test Results

performed in the initial design phase, AI research could begin to benefit onboard software without violating its constraints.

The migration of AI technology onboard will require many changes from the methods typically used in developing spacecraft data systems. These changes will not occur without complete support of project management. Although flight software developers today require a knowledge of general hardware characteristics such as sampling delays and sensor accuracies, the calibrated models of an expert system require more in-depth knowledge of the hardware. Acquiring adequate data requires the commitment of the hardware designers and test engineers, as well as project managers who must assure that sufficient time is allocated on the hardware test system. As was evident during our research, many diverse tasks must be performed on the equipment, and priorities must be established to efficiently use this limited resource. The problem exists today and will only increase with the added competition for use of the hardware by expert system developers.

CONCLUSIONS

AI's migration onboard may not be accomplished through one giant leap, but rather in a series of small steps. As spacecraft systems and missions become more complex and demand greater autonomy, more high-level control functions will be carried out by the onboard software. AI research is seeking to automate many of the high-level applications such as diagnosis and planning that are now performed on the ground. In our research we have shown that a model-based expert system can interact with prototype flight software and correctly identify several hardware faults. However, this system and other large scale AI prototypes cannot yet meet the verification and validation requirements and other restrictions imposed by the highly constrained onboard environment.

Eventually researchers will be able to verify their AI systems and flight processors will grow sufficiently in size and throughput to accommodate large expert systems. However, even when feasible, porting these full-scale AI systems onboard may not be the most effective application of this technology. The loosely coupled architecture developed for our application did not allow the programs to take full advantage of each other's capabilities and led to redundant processing. A tighter integration of AI and conventional software in which each technique is applied to those system requirements to which it is best suited may be a more effective and easily managed solution. This approach would require flight software designers and AI software developers to work together to produce onboard software capable of meeting the needs of future spacecraft missions.

REFERENCES

[Cowles 90]   S. D. Cowles, L. M. Fesq, A. Stephan, and E. J. Wild, "Expert Systems: A New Approach to Spacecraft Autonomy," TRW Quest Magazine, Volume 13, Number 1, Summer 1990.

[Davis 88]   R. Davis and W. Hamscher, "Model-based Reasoning: Troubleshooting," Proceedings of AAAI-88, St. Paul, Minnesota, 1988.

[Fesq 89]   L. M. Fesq and A. Stephan, "Advances in Spacecraft Autonomy using Artificial Intelligence Techniques," Proceedings of the Guidance and Control Conference, Keystone, Colorado, Volume 68, 1989.

[Fesq 91]   L. M. Fesq, A. Stephan, E. R. Martin, and M. G. leRutte, "Model-based Reasoning for Space Station Freedom," Proceedings of the 26th Intersociety Energy Conversion Engineering Conference, Boston, Massachusetts, 1991.

[Filarey 90]   A. R. Filarey, "Developing Onboard Spacecraft Software in Ada," Proceedings of the 1990 TRW Conference on Evolutionary Development Experience and Supporting Ada Technologies, Dominguez Hills, California, 1990

Page intentionally left blank

FIDEX: An Expert System for
Satellite Diagnostics

John Durkin, Donald Tallo
The University of Akron
Akron, Ohio

Edward Petrik
NASA Lewis Research Center
Cleveland, Ohio

ABSTRACT

A Fault Isolation and Diagnosis Expert system (FIDEX) was developed for communication satellite diagnostics. It was designed specifically for the recently completed 30/20-gigahertz satellite transponder, developed at NASA Lewis as part of the ACTS (Advanced Communication Technology Satellite) System. The expert system was designed with a generic structure and features that make it applicable to other types of space systems.

FIDEX is a frame-based system that enjoys many of the inherent frame-base features, such as inheritance, message passing, etc. The frame architecture integrates a frame hierarchy that describes the transponder's components, with other hierarchies that provide structural and fault information about the transponder. This architecture provides a flexible diagnostic structure and enhances maintenance of the system.

FIDEX also includes an inexact reasoning technique and a primitive learning ability. Inexact reasoning was an important feature for this system due to the sparse number of sensors available to provide information on the transponder's performance. FIDEX can determine the most likely faulted component under the constraint of limited information. FIDEX learns about the most likely faults in the transponder by keeping a record of past established faults. This permits the system to search first for those faults which are most likely to occur, thus enhancing search efficiency. FIDEX also has the ability to detect anomalies in the sensors that provide information on the transponder's performance. This ability is used to first rule out simple sensor malfunctions.

1. INTRODUCTION

The satellite network of the United States represents a strategic resource for this country. It supports both the commercial and military sectors by providing effective world-wide communications. The reliable operation of each satellite represents a critical goal of NASA.

Satellite reliability is presently maintained through human intervention. When a problem occurs, ground personnel are first made aware of it when the satellite communicates its status to them during a fly-by. They then use telemetry from the satellite to aid them in correcting the fault. This process proceeds through the tasks of: fault isolation, fault diagnosis and fault response. Findings are also recorded for future reference in the event similar conditions reoccur.

Since the mid 80's, NASA has investigated the application of expert system technology to replicate the satellite diagnostic tasks performed by the ground

143

personnel. The principle motivation for this work has been to develop an expert
system that can be placed onboard the satellite that will permit the satellite to
autonomously perform self diagnosis. Success in this effort offers the potential of
improved reliability in situations where ground personnel are not in communication
with the satellite quick enough to prevent its failure.

Recently, NASA Lewis completed the development of a 30/20-gigahertz satellite
transponder. The transponder is to be integrated with NASA's ACTS (Advanced
Communication Technology Satellite) System. The transponder is presently being
evaluated within the System Integration, Test, and Evaluation (SITE) system. SITE is
a laboratory used by NASA for validating designs and for evaluating and
demonstrating satellite communications systems. Figure 1 shows a diagram of the SITE
model of the ACTS transponder.
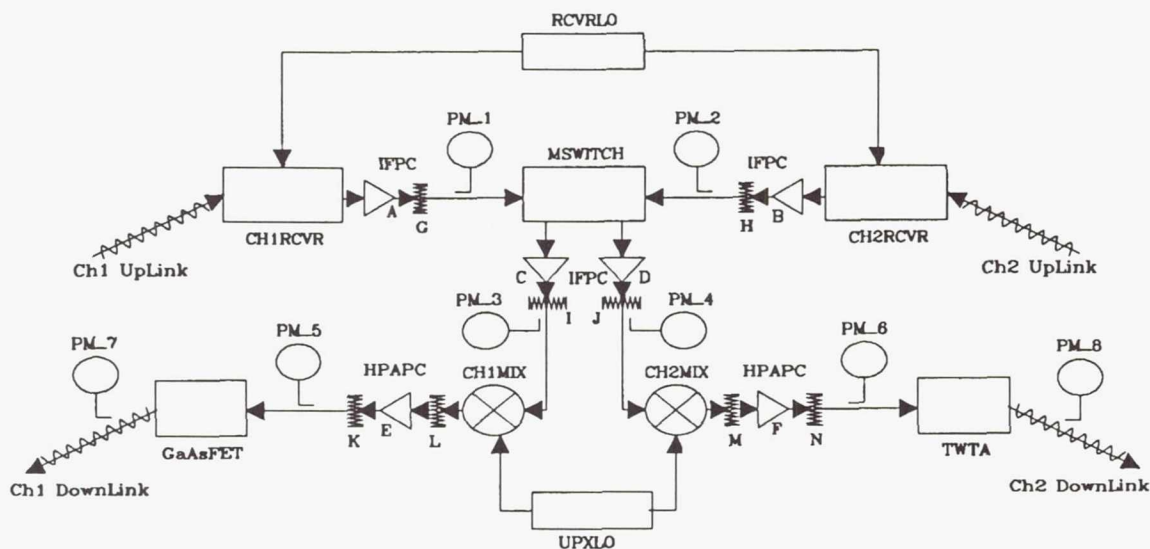
Figure 1. ACTS transponder.

Due to their interest in expert systems, NASA Lewis decided to integrate with the
development of the transponder, the design of an expert system which was capable of
performing intelligent diagnostics on the new satellite. This ongoing effort has
resulted in an expert system called FIDEX.

2. THE PROBLEM

A prerequisite for the design of most expert system projects is the existence of a
rich pool of knowledge. In a diagnostic application, this requirement usually dic-
tates that potential fault states of the system under study are well known. Since
the satellite used in this study was relatively new, the development of FIDEX had to
work under the constraint of limited diagnostic knowledge.

The transponder system is still undergoing evaluation and design changes are pos-
sible. These changes could include a modification to component designspecifications
or the addition of new components. This evolving state of the design of the trans-
ponder required that FIDEX be designed so that it could gracefullyinclude new knowl-
edge as changes are made to the transponder.

144

Another constraint placed on FIDEX was that it has to work with limited information on the operation of the transponder. The present state of the transponder has only a sparse number of sensors that provide information on the behavior of the system. Available information is limited to power levels and bit-error-rates (BER) at these few select points. The locations of the power sensors are shown in Figure 1 as PM_i.

Faced with these constraints, the work on FIDEX became more of a study effort. Techniques were developed that permitted the system to reason intelligently under the constraint of limited information. In addition, the system needed to easily incorporate changes as modifications were made to the transponder. Finally, FIDEX needed to serve as a guide to NASA for adding additional sensors to the transponder. That is, if we could demonstrate that information presently not available on the transponder's performance could be of value to the expert system, then we could make recommendations for the addition of new sensors that could provide this information. All of these requirements placed a premium on designing a knowledge representation technique and reasoning method that were general and flexible.

3. FIDEX DEVELOPMENT ENVIRONMENT

Since FIDEX needed to be designed in a fashion that would allow it to easily incorporate changes to the transponder, a frame-based approach was taken for knowledge representation. The system was developed on an IBM Model 80 PC using NEXPERT from Neuron Data.

NEXPERT permits an object-oriented style of programming within class/subclass/object hierarchies. It includes message passing through active facets and general rules that can scan the frame hierarchies. It also permits access to database information contained in dBASE III and can execute external C-language programs. In addition, NEXPERT runs in Windows 3.0 and supports dynamic-data-exchange. All of these features of NEXPERT were important in the design of FIDEX as explained in the next several sections.

4. FIDEX ARCHITECTURE

Figure 2 shows a block diagram of FIDEX. The following sections describe each of the blocks illustrated in this figure.
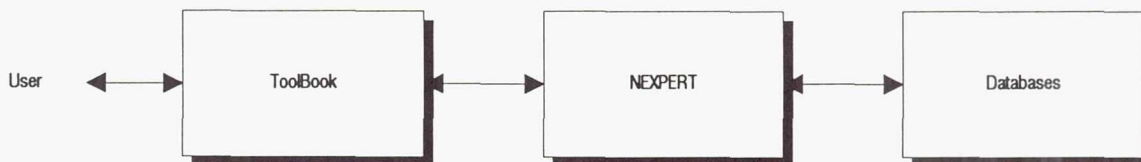
Figure 2. FIDEX block diagram.

4.1 INTERFACE

The long term objective of FIDEX is to permit it to acquire data on the operation of the transponder from a data acquisition system. However, during the development of FIDEX, it was decided to acquire this data interactively from the user through the interface package ToolBook. ToolBook runs in Windows 3.0 and, through dynamic-data-exchange, it can interact with NEXPERT.

145

The interface is highly menu driven. The user enters information about the condition of the transponder via various forms and prompts. This data is then dynamically transferred to the NEXPERT application where it is evaluated. The interface also allows NEXPERT to prompt the user for information as it is required during the diagnostic process. The results of the evaluation are transferred back to ToolBook where they are reported to the user. These results are conveyed to the user via color changes on interface diagrams and various report forms.

Figure 3 shows an example of the FIDEX interface. The main menu is displayed as the menu bar across the top of the screen. Clicking with the mouse on one of these menu topics displays a pulldown menu for that topic. The pulldown menu for sensor data input is shown that allows several options. First, all the sensors can be initialized to their nominal values by selecting "Nominal" from this menu. The user can also enter sensor data by selecting either "Form" or "Individual." Form input allows the user to input all sensor information via one form. Individual input allows the user to individually alter a sensor value.
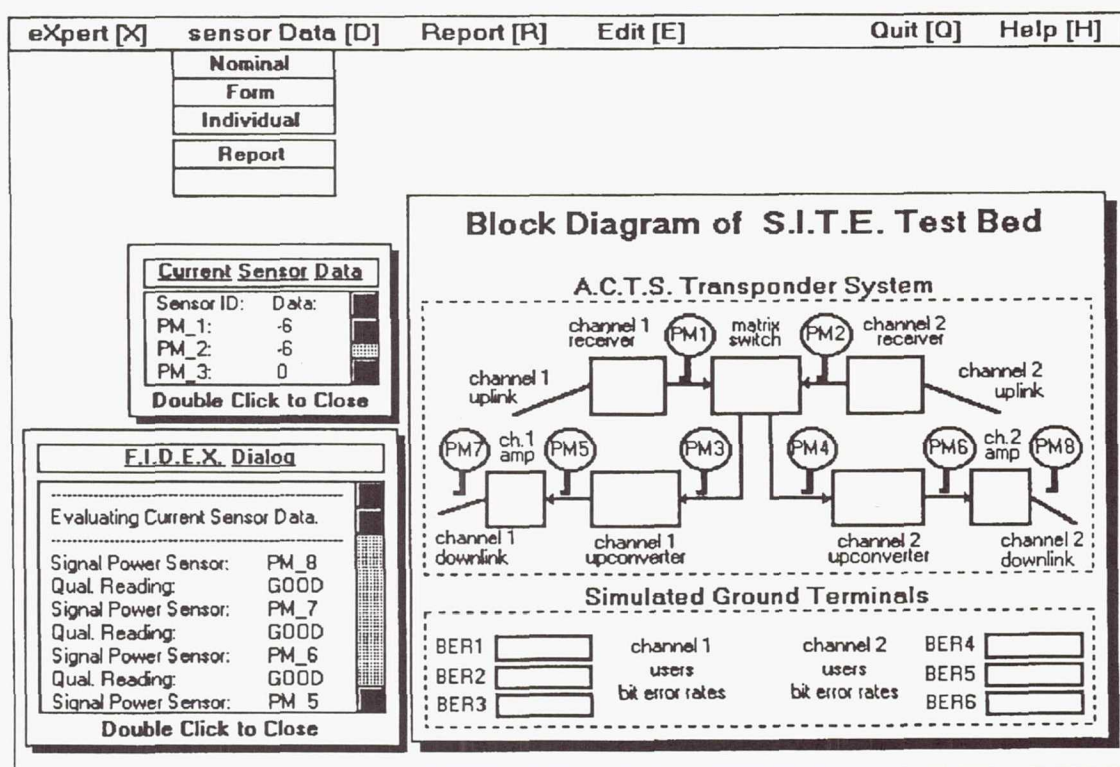
Figure 3. FIDEX interface.

The block diagram of Figure 3 shows the sensors and subsystems of the transponder. This diagram graphically displays the results of FIDEX. For example, if the fault is isolated to a subsystem, FIDEX displays this event by changing the color of this subsystem on the diagram. Also shown on Figure 3 are report forms which display sensor information and the evaluation by FIDEX on the operation of the transponder.

4.2 DATABASE

There are two databases used by FIDEX. One contains information required to initialize the sensors. Each record of this database contains information on the sensor's nominal reading, error tolerances, and other initial parameters. These values are

loaded and stored in the appropriate slots of the sensor objects. This method of intialization was chosen to facilitate system maintenance. The second database is used to provide FIDEX a limited learning capability. FIDEX stores the failure history of the transponder system in this database. Each known fault state of the transponder is represented by a record that contains a field which represents the history of that fault state. Following a session with FIDEX, the identified fault has its field value incremented. This recordkeeping is used in future sessions to direct the search towards the most likely faults.

4.3 KNOWLEDGE BASE

FIDEX's knowledge is represented in both frames and rules. Frame hierarchies were developed to represent the transponder's components, subsystems, sensors and faults. These hierarchies were also interconnected in network form to enrich the overall knowledge representation structure. The rules were written to scan the frames and were responsible for fault diagnosis. The following sections describe the frame architecture.

4.3.1 COMPONENTS WORLD

The design of the architecture for the frames used in FIDEX had to first provide a clear and efficient representation of all of the components used in the transponder. This was accomplished using the hierarchical design illustrated in Figure 4, where classes are drawn as circles and objects as triangles. The root node of Figure 4 is a class frame called COMPONENTS that contains properties common to all the children frames shown below it. The children inherit properties, values and methods from the COMPONENTS class. Also, each subclass frame has additional properties that are specific to its name and are inherited by their children. As common to any frame-base system, this structure accommodates the addition of new components as they are added to the design of the transponder.
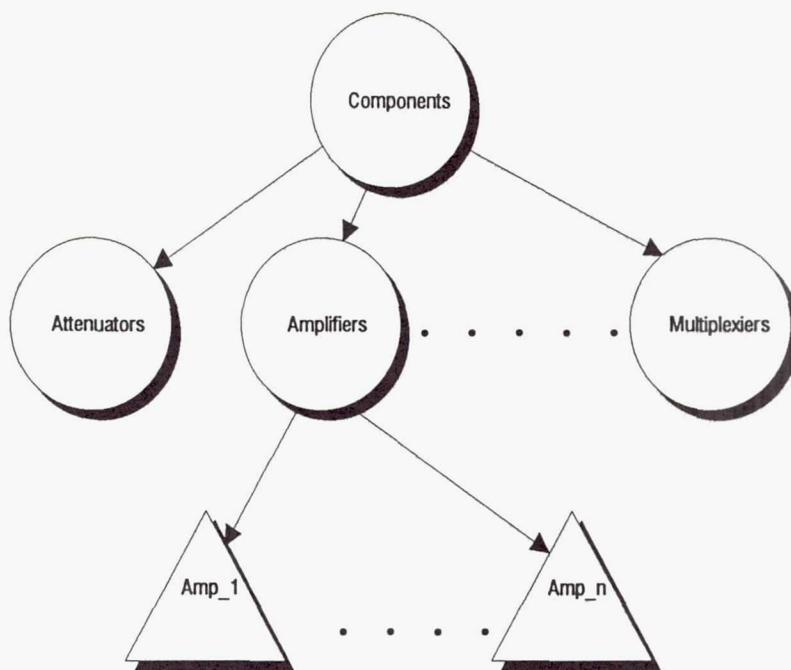
Figure 4. Components frame architecture.

147

4.3.2 SUBSYSTEMS WORLD

During system diagnosis, one of the first tasks of FIDEX is to isolate the problem to a small set of potentially faulty components. This approach enhances the efficiency of system diagnosis. To accommodate this task, the transponder system of Figure 1 was represented as several interconnected blocks or subsystems. Each subsystem has several different types of components, i.e. amplifier, attenuator, etc. Each of these types of components are represented in FIDEX as previously shown in Figure 4. Therefore, in the representation of the various subsystems, a network was formed that interconnected the world of components with the world of subsystems as illustrated in Figure 5.
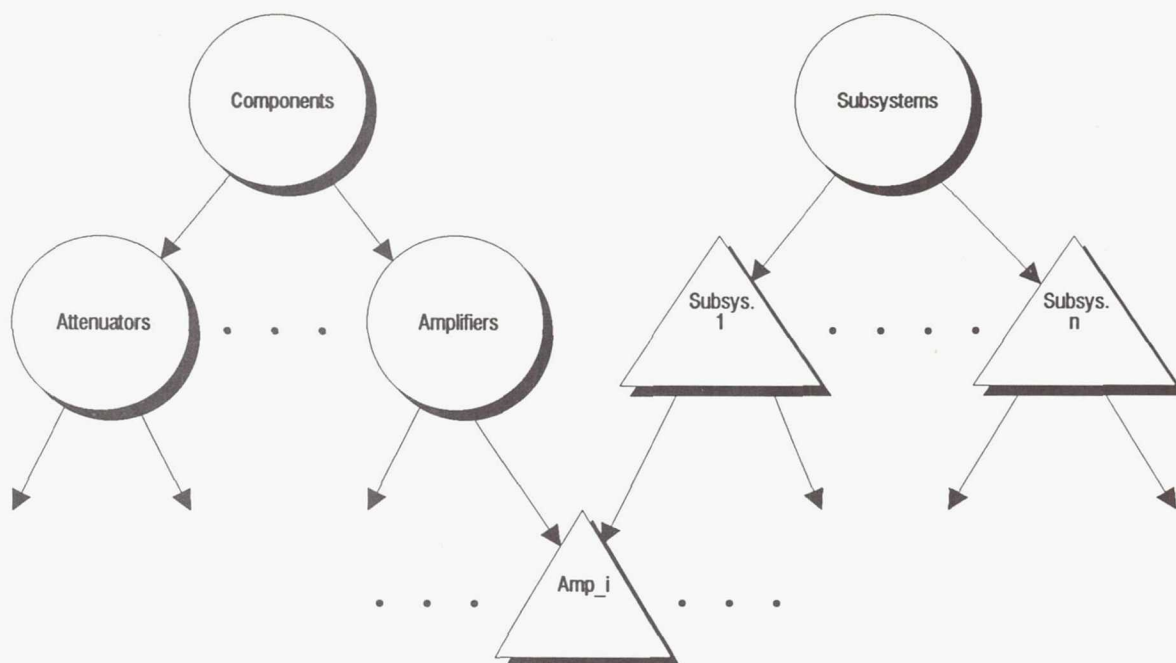
Figure 5. Subsystems frame architecture.

With this architecture, each object frame is associated with two worlds: the components of the transponder and the subsystems of the transponder. The link to the components world can be interpreted as an IS-A link while to the subsystems world as a PART-OF link. This approach not only aids the diagnostic task discussed later, but provides an efficient coding approach where each subsystem component inherits, through multiple inheritance, information from two parents - one provides information on performance while the other on structure.

4.3.3 SENSORS WORLD

Fourteen sensors monitor the operation of the transponder and the relayed signal. Eight of these are power level sensors that report the signal power levels at key

locations within the transponder system. The remaining six sensors are BER registers and are located within the ground terminal systems. They report the error, in percentages, incurred when the signal is relayed through the system. Information provided from both the power and BER sensors is used for transponder diagnosis.

FIDEX considers sensors like all other transponder components, a component that could potentially fail. It validates each sensors reading before proceeding to transponder diagnosis. Therefore, each sensor is represented in FIDEX as a member of both the sensors world and the world of components. The frame structure used to represent the sensors in FIDEX is illustrated in Figure 6.
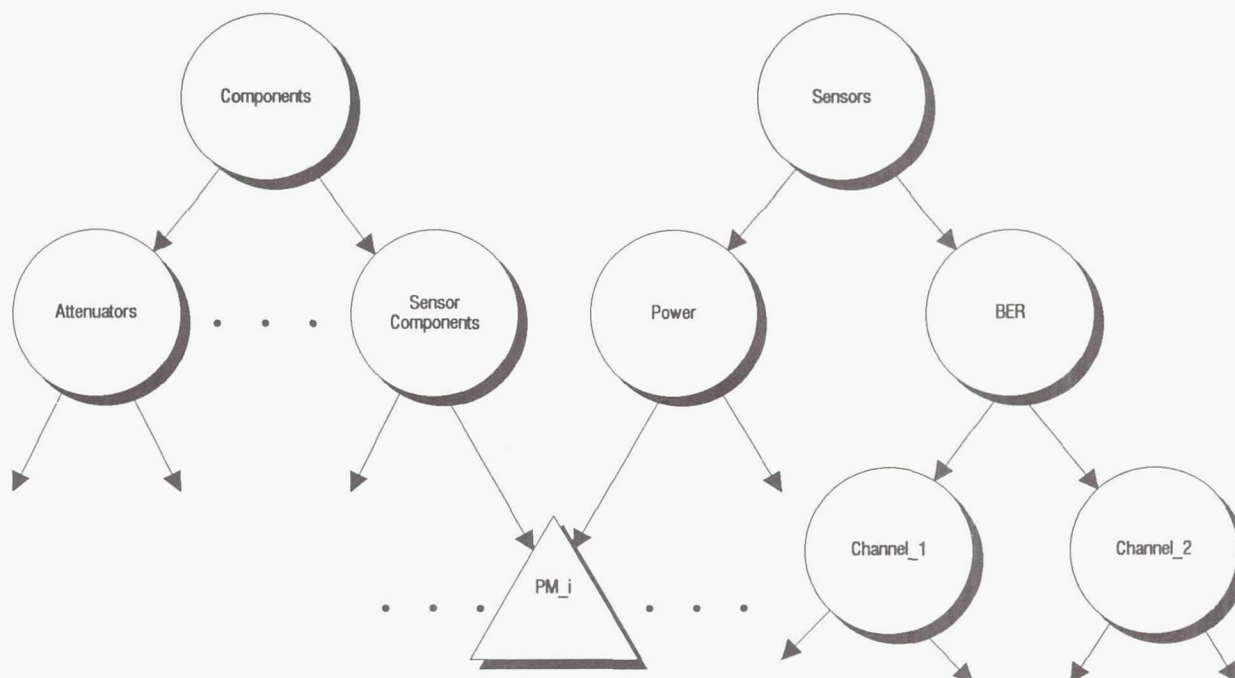
Figure 6. Sensors frame architecture.

4.3.4 FAULT STATES WORLD

The potential transponder fault states were represented in the frame structure shown in Figure 7. Objects to represent each known fault state in the transponder system are attached to nodes under the class of FAULT STATES. These nodes are used to associate the fault state objects with a type of component. For example, fault states which are associated with amplifiers are attached to the AMPLIFIER FAULTS class node.
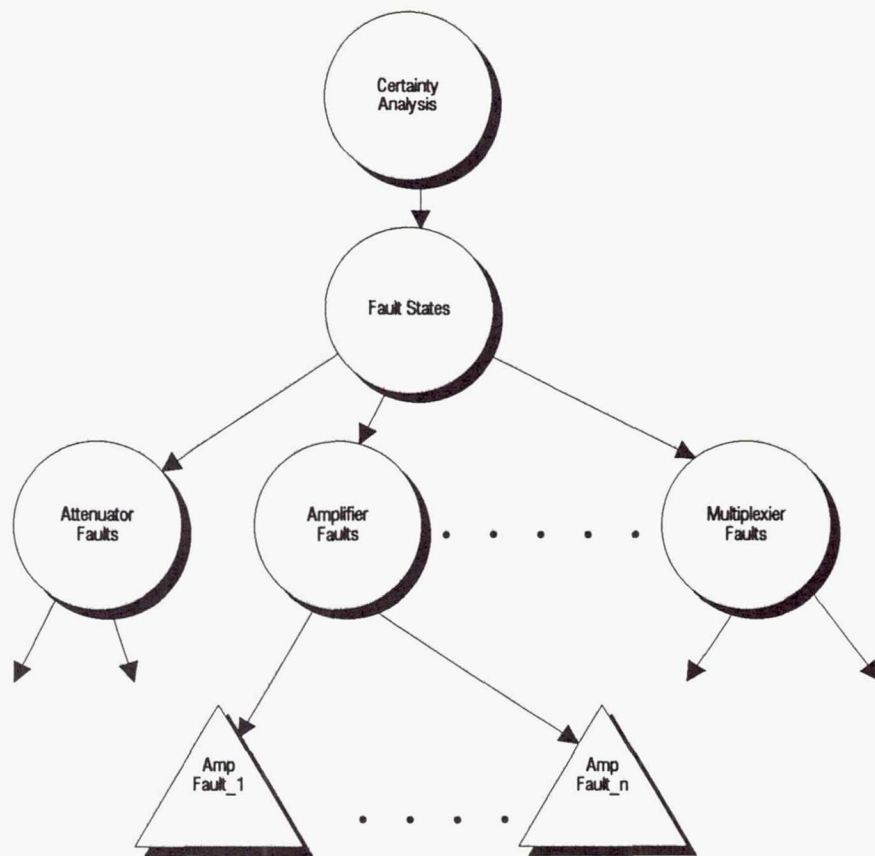
Figure 7. Fault state frame architecture.

The fault state objects of Figure 7 are generic. They can apply to any component that comes from a given class. For example, if FIDEX was considering potential faults of some amplifier, it would consider the same issues regardless of which subsystem it was a component of. This feature offers efficient coding and also permits FIDEX to easily adapt to the addition of new components to the transponder.

5. PROBLEM-SOLVING APPROACH

The problem-solving approach used by FIDEX follows that used by ground personnel who perform satellite diagnostics: fault detection, fault isolation, fault diagnosis and fault response. FIDEX performs each of these tasks using different rule modules. The sequence of tasks performed are discussed in the following sections.

5.1 TASK 1 - FAULT DETECTION

The purpose of fault detection is to detect any misbehavior in the transponder performance. This task is accomplished by a rule module that continually scans, in a data-driven fashion, the sensor frames which maintain information on the current sensor readings. Fault detection is based on a current reading exceeding a tolerance figure centered on a nominal or expected sensor value. Each sensor frame contains slots for these values. Rules ascribe a qualitative description of each sensor's reading as either GOOD or BAD, depending on whether the current reading is within tolerance. A BAD reading indicates a fault and initiates fault isolation to begin.

150

5.2 TASK 2 - FAULT ISOLATION

The fault isolation task isolates the suspected fault to one of the transponder's subsystems. This is accomplished by another rule module that considers the qualitative description of all of the signal data contain in the sensor frames. These rules locate a sensor reporting a "GOOD" reading followed by one with a "BAD" reading. The subsystem located between these two sensors is then labelled as faulty.

5.2.1 SENSOR VALIDATION

FIDEX was designed with the ability to identify a faulty sensor. This ability permits the system to avoid the search for a non-existing transponder fault. It could also be used for a reconfigurration of sensors, where faulty ones are removed. Sensor validation is based on simple error propagation. A signal producing a "BAD" sensor reading at one point in the transponder, should result in "BAD" readings in sensors measuring signals dependent on the first signal. FIDEX identifies a faulted sensor if a "GOOD" reading instead is found.

5.3 TASK 3 - FAULT DIAGNOSIS

FIDEX maintains a library of diagnostic rule modules. Each module is designed to address problems with each subsystem within the transponder. Following the isolation task, where the suspected faulted subsystem has been identified, FIDEX loads the appropriate rule module and begins to diagnose the subsystem. Each of the rule-sets perform the diagnosis using a backward chaining approach. The goals for the chosen set represent potential faults for the corresponding subsystem. They are placed on an agenda and pursued exhaustively. The order in which these goals are placed on the agenda is based on the history of the fault states which is maintained in a database. This history is used to order the goals on the agenda. This approach permits FIDEX to pursue the most likely problems first.

5.3.1 INEXACT REASONING

Since one of the constraints that FIDEX needed to work under was limited information on the operation of the transponder, it was designed with the capability to perform inexact reasoning. FIDEX uses an inexact reasoning technique based on the certainty theory (Shortliffe 1975), with some small modification. This technique relies upon establishing a measure of belief (MB) or a measure of disbelief (MD) in a rule's conclusion (H). These two factors can be used to incrementally establish an overall belief or confidence factor (CF) value for H supported by multiple rules through the use of the following equations:

$$MB(H)_{new} = MB(H)_{old} + MB(H)_{new} (1 - MB(H)_{old}) \qquad (1)$$

$$MD(H)_{new} = MD(H)_{old} + MD(H)_{new} (1 - MD(H)_{old}) \qquad (2)$$

$$CF(H) = \frac{MB(H)_{new} - MD(H)_{new}}{1 - MIN \{MB,MD\}} \qquad (3)$$

MB and MD are numeric terms that range from 0 to 1. The CF term ranges from -1 (definitely false) to +1 (definitely true). Values between these two limits represent a degree of disbelief (negative values) of belief (positive values). The

151

term $MB(H)_{old}$ $(MD(H)_{old})$ is the measure of belief (measure of disbelief) established in H from the firing of previous rules. When a new rule fires, it establishes either a $MB(H)_{new}$, if the evidence supports H, or a $MD(H)_{new}$, if the evidence rejects H. This new rule firing is also used to incrementally add to the belief or disbelief in H according to the above equations. If the evidence is in support (rejection) of H, then equation 1 (equation 2) is used. Finally, the overall belief in H is established by equation 3. These equations were embedded in the CERTAINTY ANALYSIS root frame of Figure 7 to permit their inheritance by the lower level frames.

Using this approach, FIDEX can use each piece of available evidence obtained either from sensor data or supplied from the user to incrementally add to the belief of disbelief of each fault state of Figure 7. It also permits FIDEX to conclude that an "abstract fault" exists even if it is unable to determine a "specific fault." For example, it might conclude "I believe that there is an amplifier problem in subsystem_3," instead of "Amp_1 in subsystem_3 has a bad output stage." The following rule illustrates how FIDEX can add to MB or MD of either an object level or class level fault state:

IF	The_Fault_Has_Symptoms_In_The_Frequency_Response
AND	{|CH1_BER_SENSORS|}.READING Is "BAD"
THEN	The_LO_May_Be_Out_Of_Phase_Lock
AND	Let Internal_LO_Phase_Lock_Fault.MB = 0.7 ————> specific fault
AND	Let |ATTENUATOR_FAULTS|.MD = 0.7 ————> abstract fault
AND	Let |LO_FAULTS|.MB = 0.5 ————> abstract fault

Given only one piece of evidence, this rule can establish a belief of disbelief in both specific or class level faults. The firing of this rule would also cause an incremental update in the belief of these fault states following the above equations.

5.4 TASK 4 - FAULT RESPONSE

At present, FIDEX performs fault response by providing recommendations on component or sensor reconfigurration. Future plans are to include the capability to reconsider fault diagnosis in the event the recommended action was ineffective. The system will retain its past diagnosis, including recommendations, and reconsider the problem with information made available following the reconfigurration of the transponder.

6. SUMMARY

FIDEX is an expert system designed to perform fault diagnostics on a new satellite developed by NASA Lewis Research Center. It was built with maximum flexibility, both in terms of its knowledge representation architecture and problem-solving approach, in order to adapt easily to changes to the satellite design. It was also designed in a fashion that its performance would naturally improve as performance information became available. The resultant design should be applicable to the diagnostics of other spacecraft systems, where design and performance issues are evolving factors.

REFERENCES

Shortliffe, Edward H. and Bruce G. Buchanan, "A Model of Inexact Reasoning in Medicine," Mathematical Biosciences, Vol.23, 1975.

FAULT-TOLERANCE TECHNIQUES FOR HIGH-SPEED FIBER-OPTIC NETWORKS

John DeRuiter*
Honeywell Inc.
Glendale, Arizona

ABSTRACT

Four fiber-optic network topologies (linear bus, ring, central star, and distributed star) are discussed relative to their application to high data throughput, fault-tolerant networks. The topologies are also examined in terms of redundancy and the need to provide for single-point, failure-free (or better) system operation.

Linear bus topology, although traditionally the method of choice for wire systems, presents implementation problems when larger fiber-optic systems are considered. Ring topology works well for high-speed systems when coupled with a token-passing protocol, but it requires a significant increase in protocol complexity to manage system reconfiguration due to ring and node failures. Star topologies offer a natural fault tolerance, without added protocol complexity, while still providing high data throughput capability.

INTRODUCTION

Traditionally, networks for the commercial market have been designed to provide fault tolerance. This fault tolerance, however, has only been provided to a limited extent. That is, a single fault cannot interrupt communications to all nodes but may be allowed to cause the interruption of communications to a single node or a group of nodes. This is less than desirable for aircraft and space applications where there may be critical communications between individual nodes, requiring the total system, not just the network, to be free of single-point failures. Some applications require greater than single-point failure tolerance. For example, the Space Station is required to be operational after two faults. It is therefore desirable that a network design use modular fault-tolerant techniques that can be expanded to greater levels of fault tolerance. As opposed to a commercial office environment, high-speed aerospace applications may require very rapid fault recovery to avoid data loss or excessive delays. Ideally, a network should support autonomous fault-recovery with the fault recovery mechanisms distributed at the individual nodes to provide as rapid a recovery as possible and to avoid centralized system vulnerability.

Many network architectures have been created, based on various fiber-optic-compatible topologies for both commercial and aerospace applications. Commercial systems are characterized by long runs, a relatively large number of nodes, low cost, and limited fault tolerance; aerospace systems, however, are characterized by short runs, a smaller number of nodes, low power, high reliability, and more extensive fault tolerance. This paper examines various fiber-optic topologies, their protocols relative to fault tolerance, and their applicability to the aerospace environment. First, the linear bus topology is discussed; then the ring architecture is examined. Finally, star topologies are addressed.

LINEAR BUS TOPOLOGY

Traditionally, linear bus has been the topology of choice for wire systems. Because of implementation issues, however, it has only limited applicability to fiber-optic networks. In general, the number of nodes that can be supported by a linear bus, without repeaters, is severely limited due to cascaded optical coupler/connector losses and receiver dynamic range and sensitivity limitations.

Since most optical tee couplers are unidirectional devices in which splitting ratios are not reciprocal, a linear bus topology is usually configured with separate couplers and fiber for transmit and receive,[1] as shown in Figure 1. It can be seen that if all transmitters have the same power output and all couplers have the same splitting ratio, the dynamic range requirements imposed on the receiver will increase as the number of nodes in the network increases. In addition, because of the cumulative effect on attenuation of cascading couplers and connectors, receiver sensitivity requirements also increase in proportion to the number of nodes. When considering LED transmitter power, coupler/connector loss, and dynamic range/sensitivity characteristics of available pin

*Staff Engineer; Honeywell Inc., Satellite Systems Operation, Glendale, AZ

diode receivers at high data rates, network size is limited to approximately six nodes if coupler splitting ratios and transmitter powers are fixed. Varying the transmitter power or coupler splitting ratios decreases the dynamic range requirements on the receiver but does not substantially decrease the receiver sensitivity requirements. Accumulated connector, fiber, and coupler losses prevent the linear bus from supporting much greater than seven nodes, even when techniques are used that limit the dynamic range requirements.
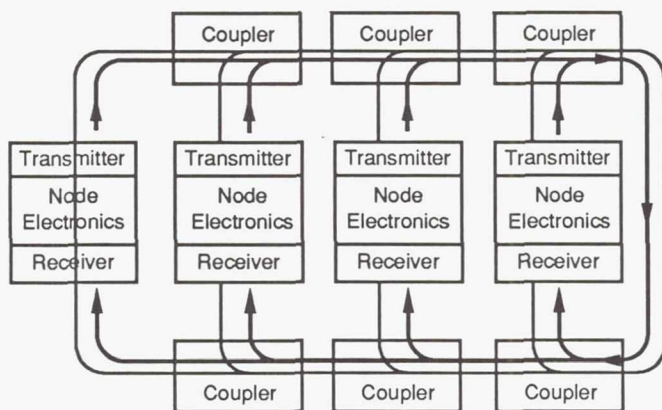
Figure 1. Linear Bus Topology

Linear Bus Fault Tolerance

Fault tolerance for the linear bus requires a duplication of both fiber and couplers. Figure 1 shows that if only a single fiber fails between either a receiver or transmitter and a coupler, only a single node will be affected. If, however, a coupler fails, total network failure can result. It is, therefore, necessary to duplicate all couplers, fibers, and node electronics (as shown in Figure 2) to provide a single fault-tolerant network. This technique can be extended if greater fault tolerance is desired. Should a failure occur in the primary network, however, all activity must be switched to the backup network leaving serviceable node electronics idle. This can be remedied by cross-strapping the primary and
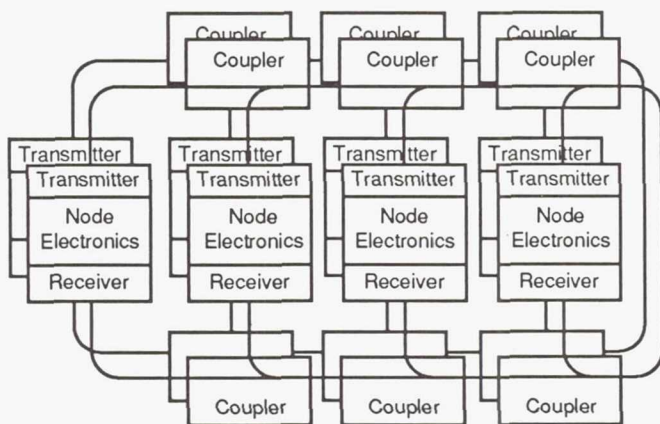
Figure 2. Redundant Linear Bus Topology

backup node electronics to both primary and backup networks. Unfortunately, this adds attenuation to the linear bus, further decreasing the already limited number of nodes that can be supported. The linear bus appears to be less than ideal when applied to any reasonable-size, fiber-optic network with fault-tolerant requirements.

RING TOPOLOGY

Basic ring topology, shown in Figure 3, has the advantage of being a group of point-to-point links, with each node being an active repeater; thus it requires no optical couplers. The dynamic range and sensitivity problems that limit the number of nodes in a linear bus topology are, therefore, substantially eliminated with the ring. Unfortunately, the network is now subject to total failure if any single node or fiber fails. Redundant components can be used to overcome this problem. Interruption of network communication is now, however, a function of active components (repeaters) as opposed to passive components (optical couplers).
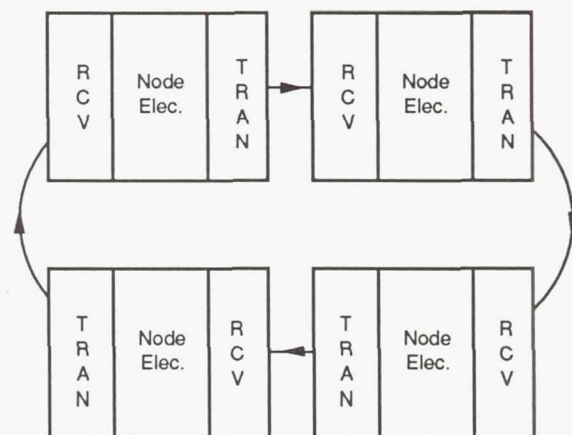
Figure 3. Basic Ring Topology

Ring Protocol

To provide for deterministic operation and high efficiency at high data rates, a token-passing protocol is typically used on a ring. The token ring protocol is based on the idea of a free token circulating around the ring. When a node desires to transmit, it captures the token and then transmits its data. Upon completion of its transmission, the token is reissued. Subsequent stations on the ring then have the opportunity to capture the token and to transmit their own data. Additionally, these token protocols incorporate features to recover from errors on the ring that cause total disruption of network communication (in particular, lost tokens due to bit error rate effects). This is done, however, at the expense of protocol complexity and ring down time. For example, an FDDI system, upon detection of a lost token, requires that all nodes enter a "claim token" mode and, in concert, determine which node has the highest priority and,

154

therefore, the right to transmit and issue a new token. This increases protocol complexity, and, due to the needed cooperation between all nodes, necessitates an interruption in communication to all nodes.

Ring Fault Tolerance

To provide fault tolerance within an optical ring topology, and not just to accommodate soft-error recovery, two additional techniques can be used, including optical bypassing and counter-rotating rings.

A failed node can be bypassed using an optical switch. In a spacecraft application, where power and reliability are critical, it is advantageous to power down any unused nodes both to lower power and to increase reliability. The optical bypass provides a means to circumvent these powered-down nodes. Bypass control can be a completely distributed function, with each node providing autonomous fault detection and bypass. Unfortunately, the optical bypass switch adds attenuation between nodes and, together with optical receiver sensitivity and dynamic range capabilities, limits the number of adjacent nodes that can be bypassed. Only about three adjacent nodes can be bypassed. This is a small number considering a ring's capability of supporting a large number of nodes. In systems where it is desirable to power down a large number of nodes to decrease power consumption and to increase reliability, the ring limits flexibility because care must be taken in how many adjacent nodes are powered down. An additional consideration is that ring operation is interrupted for a finite amount of time because of the bypass switching time. This time can be as great as 25 milliseconds. For high-speed networks, relatively large queues can be required within the node electronics to prevent data loss due to the network communication disruption, which is caused by bypass switching time.

Whereas the optical bypass provides a means for bypassing powered-down or failed nodes, the counter-rotating ring provides for proper ring operation even after a fiber cable has failed. Figure 4 shows how the ring would reconfigure if a cable break should occur. Even though there is a cable break, all nodes can still communicate over a ring that is approximately twice as long as the original. This increases ring latency, but, for aerospace application where run length is relatively short, this effect is insignificant. This provides for single fault-tolerant operation on a system basis if each node is internally dual redundant or if redundant nodes are inserted into the ring. Ring reconfiguration is accomplished by a cooperative effort between all nodes on the ring to locate the break, initiate the necessary reconfiguration, and reinitialize the network. The expense of this cooperative effort, as with recovery from a lost token, is an increase in the network protocol complexity and, as with the bypass, a temporary interruption of network services to all nodes.
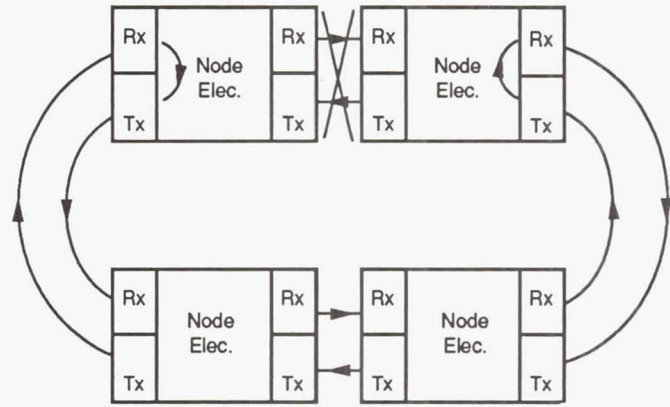
Figure 4. Reconfigurated Counter-rotating Ring (Cable Failure)

The ability of the ring topology to satisfy greater than single fault-tolerant requirements is not a simple extension of the counter-rotating ring technique. It can be solved, however, by adding additional rings. This, like the redundant linear bus, requires that all activity be switched to the backup network. This does not provide the optimum reliability, since it leaves serviceable node electronics idle on the failed ring or rings. Cross-strapping can be implemented to solve this problem, as shown in Figure 5, effectively providing active nodes as a bypass mechanism instead of a simple optical bypass. It is still desirable, however, to incorporate optical bypasses to allow for power down of all node electronics. If a node fails, the network first goes into "claim token" mode, and then into "beacon" mode to identify the failed node. The network manager can then issue a command to the backup node electronics to insert itself into the ring. In this manner, serviceable node electronics are conserved. Unfortunately, the total system is affected by this reconfiguration, not just the failed node, thus incurring an interruption in services to all nodes.
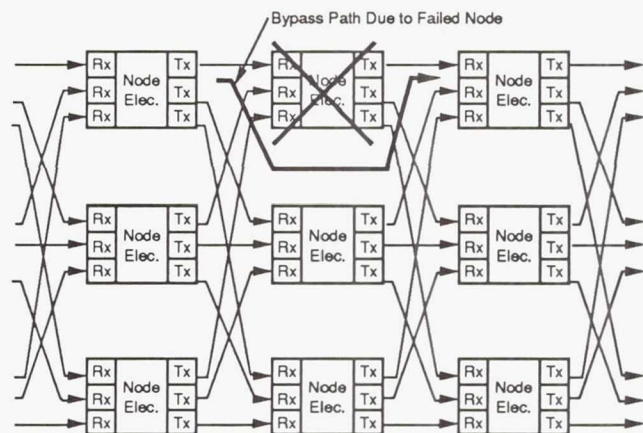
Figure 5. Two-fault-tolerant Cross-strapping for Ring Topology

155

STAR TOPOLOGIES

Star topologies offer another, and in many cases, better choice for high-speed, fiber-optic, fault-tolerant networks. As shown in Figure 6, a centralized star topology is composed of a variable number of nodes interfaced via a star coupler. This star coupler can be either active or passive. Considering, however, the high reliability requirements of the desired networks, only passive optical star couplers are considered here because of their greater reliability. The star topology, like the ring, overcomes the need for optical receivers with large dynamic ranges. Unlike the ring, however, as the number of required nodes in the network increases, so does the attenuation in the star coupler. This requires greater receiver sensitivity or higher transmitter power for larger networks. Using LED emitters and PIN diode receivers, the star topology can support networks of 50 nodes, which should be quite sufficient for most aerospace applications. Up to 200 nodes can be supported by making use of laser diodes and avalanche photodiodes. Unlike the ring topology, however, the star is not susceptible to total network failure or disruption due to the failure of a single node or fiber. It can also incorporate cross-strapping techniques in its redundant configurations that make more efficient use of system components without added protocol complexity and, therefore, improve both system reliability and fault tolerance. Another advantage of the star topology is that no bypass mechanisms are needed at powered-down nodes, which allows any number of nodes or any sequence of nodes to be powered down. This offers potential power savings and better reliability for those systems that have requirements for a "sleep" mode where a large percentage of the nodes are inactive or not used.
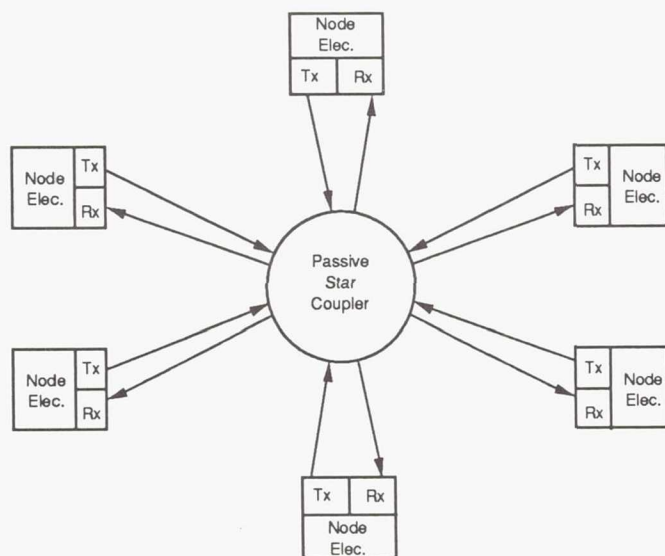
Figure 6. Basic Star Topology

Star Protocol

Both token-passing and contention-type protocols can be implemented on a star topology. At low data rates, 10Mb/s and below, both are efficient.[1] At high data rates, token passing becomes inefficient because of the greater token-passing overhead associated with the star topology. Similarly, at high network loads, traditional contention protocols become inefficient because of the contention resolution algorithms that are used. Two contention-type protocols, however, offer both high efficiency and deterministic operation that make the star topology especially applicable to high-speed, fault-tolerant networks. Both of these protocols (Honeywell's Star*Bus protocol[2,3] and Network Systems' HYPERchannel™) resolve contentions in a deterministic manner via a time-slot cycle. This allows the network to have the efficiency and deterministic properties of a time-slot (virtual token) protocol and the simplicity and fault-tolerant advantages of a contention protocol. That is, since no tokens are passed from node to node, tokens cannot be lost. Fault recovery from lost tokens is, therefore, not necessary; thus, system fault tolerance is enhanced and the protocol and overall system operation are simplified.

Star Fault Tolerance

The star can be made fault tolerant through simple duplication of components. This duplication, as with the redundant linear bus or ring, requires that all activity be switched to the backup network, leaving serviceable node electronics idle. Interruption of network services to all nodes also occurs if a coupler fails due to fault detection and switching time.

Without added protocol complexity, the cross-strapping techniques shown in Figure 7 can be used with the star topology to make more efficient use of system components, while providing virtually instantaneous fault recovery in the event of a transmitter, receiver, fiber, or coupler failure. Node electronics can consist of nonredundant elements, as shown in Figure 8, or internally redundant elements, as shown in Figure 7. Nonredundant elements offer the advantages of providing a building-block approach to fault tolerance and minimal impact to the user that does not require redundancy. The use of nonredundant elements, however, requires additional taps on the star couplers: two per node for a dual system, three per node for a triple, etc. The use of internally redundant elements implies added complexity, but limits the number of taps necessary on the star coupler to one per node and, therefore, has an advantage relative to network loss budget. In both cases, the cross-strap at the optical media works the same. As shown in Figure 9, both transmitters generate identical data, with one transmitter

being interfaced to the primary coupler and the second being interfaced to a backup coupler. At the receiving node, both receivers are active, with only one being selected. If both receivers pick up a signal, priority is given to receiver A. If only one signal is present (indicating a failed transmitter, fiber, coupler, or receiver), the active receiver output will be selected. In this manner, with dual transmitters operating in parallel in the sending node and dual receivers selecting the active channel in the receiving node, virtually instantaneous fault recovery is provided. No channel selection is performed at the system level; thus overall system management is simplified. Because of the fault-tolerant nature of this configuration, override capability is provided to allow for test functions.[2]
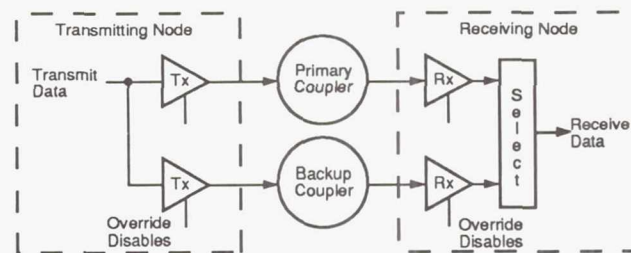
Primary Active	Backup Active	Selected Channel
1	1	Primary
1	0	Primary
0	1	Backup
0	0	Neither

Figure 9. Cross-strap Operation

Another star topology is represented by the distributed star[1] shown in Figure 10. This configuration offers greater fault tolerance than the central star in that no single optical coupler can cause the interruption of communication to all nodes. It does, however, have greater connector and excess coupler losses because of the cascading of couplers. This, in general, limits the number of nodes it can support relative to the central star approach. To achieve total system fault tolerance, the same component duplication and cross-strapping techniques (as previously described for the central star topology) can be used.

Figure 7. Dual-redundant Cross-strapped Star (With Internally Redundant Node Electronics)

Figure 10. Sixteen-node Distributed Star

AUTONOMOUS NODE FAULT RECOVERY

Some fault-recovery tasks must be performed regardless of topology. One of these tasks is switching from primary to backup node electronics. This can be done on a system basis, for example, as discussed previously for the ring topology. It is, however, more desirable to provide autonomous fault recovery at each node, thus minimizing the functions required of the network manager, not just in the ring but in all topologies.

Figure 8. Dual-redundant Cross-strapped Star

157

Figure 11 shows a detailed block diagram of a possible implementation for an autonomous switchover scheme between two nonredundant units. Each unit is identical, with the "primary ID" causing one to power up as the primary and the "backup ID" causing the other to power up as the backup. The primary is fully powered on and, therefore, fully functional. The backup is in "standby", with only its power-up control, toggle and override detectors, and receivers powered up. The backup monitors the primary's health via the toggling health signal between the two units. The CPU in the primary evaluates built-in-test results and pulses the toggle generator if all tests pass. The presence of the toggling signal is therefore the result of proper operation. The lack of a toggle from the primary will cause the primary to go into "standby" and the backup to become active and take over the node. An override command can be received, via the network, to provide for switchover testing and contingency operations in the event that a failure is not detected or detected in error. To ensure that both primary and backup are not powered simultaneously, the power switch and the override detector are made redundant. Also, the primary toggle detector will put the primary into standby if the backup powers up in error and the backup toggle becomes active. The cross-strapped toggles, therefore, provide a flip-flop type of configuration, with the primary and backup always in opposite states.

GENERAL POWER AND RELIABILITY CONSIDERATIONS

A companion issue to fault tolerance and redundancy is reliability. The intended result of providing redundancy is to enhance overall system reliability. As system components become more unreliable, greater levels of redundancy are necessary to maintain overall system reliability. Reliability is, in general, affected not only by component reliability, but also by circuit complexity and power dissipation. As circuit complexity goes up, component count goes up and reliability goes down. As power dissipation increases, component junction temperatures increase and reliability goes down. Basic differences in protocol complexity have already been discussed and are relatively clear cut. Power dissipation differences between topologies and protocols are, however, more subtle and are discussed here.

The basic nature of a ring topology, coupled with a token-passing protocol, implies a greater power usage than a star topology coupled with a broadcast protocol. In a star topology, only one transmitter is active at any one time. When a node wishes to transmit, it simply monitors the network for activity and transmits its frame when the network is free. This transmission is then received, via the star coupler, by all nodes and requires no action from other than
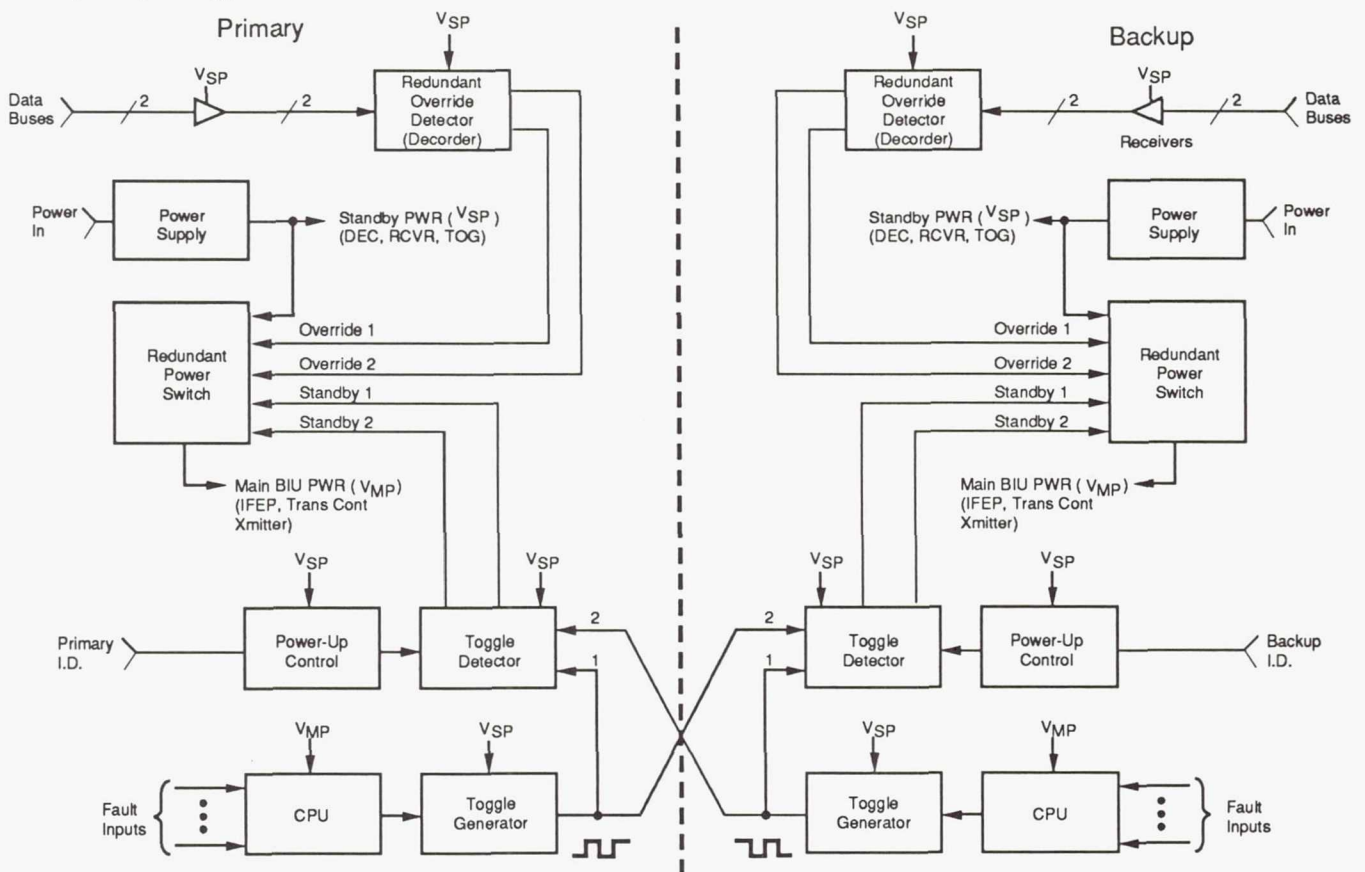
Figure 11. Switchover Implementation

the transmitting and receiving nodes. A ring, on the other hand, requires all nodes to participate in the data transfer. A frame generated by a node in a ring, however, must be repeated by all nodes on the network. This requires a greater duty cycle at each node, with each node having to transmit all frames even though they are not locally originated. In the worst case, should the physical ring be short relative to the transmitted frame, all transmitters will be active simultaneously. For example, at 100 Mb/s, a ring with a one-kilometer circumference requires all transmitters to be active simultaneously when a frame of only 500 bits or greater is circulating on the network. For aerospace applications, where runs are short, this worst-case condition is typical, not merely an exception. Each node in a ring must, therefore, run at a substantially higher duty cycle than each node in a comparable star topology; thus a higher power dissipation is incurred. This causes junction temperatures to elevate and reliability to suffer.

Other power/reliability considerations involve the protocols themselves, without regard to topology. Power-strobing techniques have long been used, in electronics intended for space applications, to reduce power dissipation, and, subsequently, to increase reliability. Circuitry is powered on only when operation is required. This makes power dissipation, and therefore reliability, dependent on duty cycle. The basic nature of a network, in which each node occupies only a portion of the network bandwidth, makes the application of power strobing beneficial. Protocols intended for use in high-reliability systems with only limited power available should be designed to allow the use of these power-strobing techniques while still maintaining high performance. Whether for a linear bus, a ring, or a star topology, the selected protocols should allow for a substantial portion of the circuitry in individual nodes to be powered off when no data transactions are occurring.

SUMMARY

Three basic topologies, relative to their application to fault-tolerant, high-speed networks for aerospace applications, have been examined. Of these three, both the ring and the star are viable candidates. The linear bus presents implementation problems for all but the smallest networks because it is limited in the number of nodes it can support. The ring can support the largest number of nodes and can easily support high data rates and deterministic operation. It can also support various levels of fault tolerance but does so at the expense of fault recovery time and an increase in media access and network management protocol complexity. The star topologies offer a better choice, providing more inherent fault tolerance, while still providing support for high data rates, deterministic operation, and a relatively large network size. The star topology also provides an inherently lower power dissipation; only one node is required to transmit a frame as opposed to the ring where all nodes must repeat the frame. Similarly, since it requires no bypasses for powered-down nodes, the star topology offers potential power savings and better reliability for those systems requiring a "sleep" mode where a significant percentage of the nodes are inactive during a particular mission phase.

REFERENCES

1. Uhlhorn, R.W., "Fiber Optic Architectures and Components for High Data Rate Buses for Military Applications," FOC/LAN 1984, pages 265-272.

2. Varga, Mike, "A Robust 100Mb/s Network for Avionics Applications," NAECON, May 23, 1985.

3. DeRuiter, John, "A Robust Fiber-Optic Network for Space Applications," OE-LAS '86, Vol. 616, January 1986, pages 214-221.

Page intentionally left blank

Fault-Tolerant Multichannel Demultiplexer Subsystems[1]

Robert Redinbo
Department of Electrical Engineering and Computer Science
University of California
Davis, CA 95616 USA

Abstract

Fault tolerance in future processing and switching communication satellites is addressed by demonstrating new methods for detecting hardware failures in the first major subsystem, the multichannel demultiplexer. An efficient method for demultiplexing frequency slotted channels employs multirate filter banks which contain fast Fourier transform processing. All numerical processing is performed at a lower rate commensurate with the small bandwidth of each baseband channel. The integrity of the demultiplexing operations is protected by using real number convolutional codes to compute comparable parity values which detect errors at the data sample level. High-rate, systematic convolutional codes produce parity values at a much reduced rate, and protection is achieved by generating parity values in two ways and comparing them. Parity values corresponding to each output channel are generated in parallel by a subsystem, operating even slower and in parallel with the demultiplexer that is virtually identical to the original structure. These parity calculations may be time-shared with the same processing resources because they are so similar.

Introduction

The new generation of sophisticated processing and switching communication satellites with their extensive digital processing capabilities will be more susceptible to temporary and permanent electronic failures. An overview of a typical satellite's subsystems is given in Figure 1. This paper will concentrate on the demultiplixer subsystem to demonstrate the general principles of fault tolerance needed in implementations that process multiple channels with shared resources. Each subsequent subsystem will have specialized features requiring variations on these fault tolerance techniques. The basic philosophy is still applicable though. Continuing work is addressing the other subsystems.

Data channels in communication systems are easily combined according to frequency division multiplexing (FDM). This method is particularly useful because frequency selectivity is all that is required to extract individual channels from the overall signal constellation. Many satellite communication systems employ this method of multiplexing since there is no requirement for common timing synchronization between data channels. This approach is even more appealing from a hardware implementation viewpoint because very efficient demultiplexer realizations, called polyphase multirate filter banks, are available [1-3].

The basic demultiplexing philosophy envisions a narrow band filter extracting each channel from the multiplexed signals. N multiplexed channels, each with relative bandwidth f_B, are combined into an FDM signal, and the corresponding demultiplexer employs an idealized narrow band filter with Z transform transfer function H(Z), shifted in frequency, to separate the band of frequencies corresponding to a channel. The uniform baseband filter is effectively shifted to each respective band by the scaling phasors, and the output channel only produces samples at a rate $1/N^{th}$ of the input sampling rate [4]. Generally, the uniform filter represented by H(Z) has a finite impulse response (FIR) configuration with the attendant advantage of linear phase [5].

The uniform filter banks can be realized by defining certain segments of the baseband filter's transfer function and then using the outputs from these shorter filters as simultaneous inputs to a discrete Fourier transform operation [1,4]. This approach to demultiplexing is outlined in Figure 2, where a fast Fourier transform (FFT) algorithm realizes the discrete Fourier transform. The relationship between the new shorter segmented filters $H^{(r)}(Z)$, r = 0, 1, ..., N-1, and the original baseband filter H(Z), will be summarized later. The most important feature of Figure 2 is its slower sampling rate applied BEFORE the filtering and FFT operations, permitting the digital hardware implementing these functions to operate at a data rate $1/N^{th}$ that of the input data sampling rate. Nevertheless, the input is still sampled at a suitably high rate commensurate with its wider bandwidth.

There are situations where this efficient form of demultiplexer must be highly reliable. Yet, the very efficient sharing of processing resources makes this form extremely sensitive to even simple failures which can easily contaminate many data channels simultaneously. For example, the new generation of switching and data processing satellites take advantage of these forms of demultiplexers.

There are several considerations when incorporating fault tolerance in such a demultiplexer system. The first

important consideration is the detection of failures, whether they are permanent or temporary and transient. Once inaccurate performance is detected, the failed subunit must be identified and located (diagnosis). Finally, if the failures persist, the system must be reconfigured so that adequate performance is still achieved. The work presented here concentrates on the first aspect, fault detection. For, without an indication of improper operation, the other aspects of fault tolerance cannot be invoked.

An emerging alternate method of fault tolerance, termed by some Algorithm-Based Fault Tolerance (ABFT), views the algorithmic operations and the data sample flow as the important items to protect regardless of the underlying hardware realization. The first use of this technique was in protecting matrix operations [9], and there have been many other applications investigated [10-17]. Most research has been directed to protecting linear algorithms.

The fundamental approach in ABFT employs real number error-detecting codes to define parity values associated with a group of data samples. These codes can be either block or convolutional codes [18-20]. In either case, the original processing algorithm is combined with the parity generation process, generally leading to a composite, efficient, simplified parity generation algorithm that produces independent parity values which are associated with the output data. Then comparable parity values are computed directly from the original processing algorithm's output data. The respective parity values, one from each set but computed in different ways, should be identical, except possibly for some small round-off error differences since they are evaluated in two dissimilar ways. Errors are detected when the respective parity values differ significantly. This type of fault tolerance will be applied to protecting the demultiplexer.

Basics of Filter Banks

An analysis bank of filters will be examined where each of the L transfer functions $H_0(Z), H_1(Z), ..., H_{L-1}(Z)$ bandlimit their respective signal outputs so that each may be sampled at a rate 1/N of the input rate. This general setting is depicted in Figure 3a. The L transfer functions will be assumed FIR types, for the purposes of the exposition. Each of the L filter paths can be analyzed separately and the generic situation is isolated in Figure 3b, for further development. The Z transform quantities shown in these figures employ the two-side Z transform. Infinite limits in the summations are included in its definition below, even though only a finite number of nonzero terms appear for the FIR filter case.

The impulse response $\{h_p(m)\}$, corresponding to the transfer function H(Z), is segmented into N subsequences and the weighting operation separated into N parallel convolutions. The N parallel convolutions employ segmented impulse responses related to the original p^{th} channel impulse response in the following way.

$$y_p(r) = \sum_{v=0}^{N-1} \left[\sum_{u=-\infty}^{+\infty} h_p^{(-v)}(r-u)\, x(uN+v) \right]$$

$$= \sum_{v=0}^{N-1} \left[\sum_{u=-\infty}^{+\infty} h_p^{(v)}(a)\, x(r-u)\, N-v \right] ; \quad (1a)$$

$$h_p^{(v)}(a) = h_p(aN+v) \quad ; \quad \begin{matrix} v=0,1,...,N-1 \\ a=0,\pm1,\pm2,... \end{matrix} \quad (1b)$$

where r in (1a) = 0, ±1, ±2, ... The same number of operations are performed in this approach, but each parallel self-contained path can operate at a rate reduced by factor N. The ability to use a slower processing rate in each disjoint parallel path provides a serial-to-parallel tradeoff.

Demultiplexers can be viewed as a special form of Figure 3a, wherein L = N and the N transfer functions are constrained to be related to one common baseband transfer function H(Z). It is easy to demonstrate the effects of the scaling phase sequence $\{e^{j\frac{f_s}{N} pr}\}_{r=-\infty}^{+\infty}$ is to shift the filter response [1]. Under these conditions, the output of the pth filter path may be written in terms of the input weighted by segments of the impulse response and scaled by a phasor.

$$y_p(r) = \sum_{v=0}^{N-1} e^{j\frac{f_s}{N} pv} \left\{ \sum_{u=-\infty}^{+\infty} h^{(-v)}(r-u)\, x(uN+v) \right\} \quad (2)$$

The final summation over index variable v in equation (2) is equivalent to forming the pth discrete Fourier transform coefficient for the N outputs of the segmented impulse response filters $\{h^{(-v)}(r)\}_{v=0}^{N-1}$. Furthermore, the only change needed to get an output for a different output, say $\{y_m(r)\}$, is to modify the scaling coefficients, $\{e^{j\frac{f_s}{N} mv}\}_{v=0}^{N-1}$, affecting the outer sum. The same segmented impulse response filters are employed, but the scaling values change. Thus Figure 2 represents the general case where an FFT form of a discrete Fourier transform is applied to the respective outputs of the segmented filter functions. All N channel outputs of the demultiplexer are obtained simultaneously.

Real Convolutional Codes

Convolutional codes have been defined traditionally over finite field alphabets [21, 22], but recent research results show how they may be extended to systems using either integer or real arithmetic [18, 20, 14]. Nevertheless, the basic approach to convolutional codes remains the same, particularly with regard to a matrix description of the

162

encoding and parity checking functions. Only systematic forms of convolutional codes will be considered primarily because the normal filtering operations are not altered and such forms are automatically noncatastrophic [22]. Only the detecting capabilities of such codes are used; any correcting operations could easily exceed the original processing requirements.

The encoding matrix for a systematic convolutional code, G, has a block-type format involving m fundamental finite sized matrices whose dimensions are related to the rate and number of parity check positions in the code. The parameter m determines the constraint length of the code.

$$
G = \begin{pmatrix}
G_0 & G_1 & - & - & G_m & 0 & - \\
0 & G_0 & - & - & G_{m-1} & G_m & - \\
0 & 0 & - & - & - & G_{m-1} & - \\
0 & 0 & | & | & - & - & | \\
0 & 0 & - & - & - & - & - \\
| & | & | & G_0 & - & - & | \\
- & - & - & | & G_0 & G_1 & - \\
| & | & | & - & 0 & G_0 & - \\
- & - & - & - & 0 & 0 & - \\
| & | & | & | & | & | & |
\end{pmatrix} \quad (3)
$$

The $k \times n$ submatrices G_j, $j = 0, 1, ..., m$, have distinctive forms and divide into two types.

$$G_0 = (I \mid P_0) \quad ; I, \; k \times k \quad \text{Identity Matrix}$$
$$P_0, \; k \times (n-k) \quad \text{Parity - Check Matrix} \quad (4a)$$

$$G_j = (0 \mid P_j) \quad ; 0, k \times k \quad \text{Zero Matrix}$$
$$; P_j, \; k \times (n-k) \quad \text{Parity - Check Matrix} \quad (4b)$$

The entries in the parity check submatrices P_i may be either 0 or 1 even for the real Marshal code case [18, 20], or in the more general case, real numbers [14, 23].

The parity positions are a function of possibly $M = (m + 1)k$ input samples through the action of the P_j parts of each G_j. The stack of these parity weighting values will be denoted by an $\{M \times (n-k)\}$ matrix Q with respective columns $\{q_r\}$. The $(n-k)$ parity position associated with the input values are obtained by the weighing action of columns q_r. Each parity value may be viewed as the output of an FIR filter, described notationally using the Z transform of column q_c:

$$Q_c(Z) = \sum_{j=0}^{M-1} q_c^{(M-1-j)} Z^{-1}, \quad (5)$$

where $c = 0, 1, 2, ..., (n - k - 1)$.

Real convolutional codes can also be imbued with a distance structure similar to the usual one applied to finite field symbol codes. It is possible to define a metric in terms of a real Hamming weight.

High rate convolutional codes with only one parity channel will be used for protecting output data channels emanating from a demultiplexer. Binary-based codes, for which there exist tables of high performance codes [24], will be chosen. In particular, a rate $K/(K+1)$ systematic convolutional code is defined by a single parity weight filter, equations (5). A single parity value for every K input sample is produced by sampling an FIR filter with transfer function denoted by $Q(Z)$. The data flow normally and are simultaneously tapped to this FIR parity filter, $Q(Z)$. A convenient view of the parity production process is shown in Figure 4. The data flow normally and are simultaneously tapped to this FIR parity filter, $Q(Z)$. The downsampling symbol $\downarrow K$ indicates that after every K data samples, one parity value is produced.

Composite Filtering and Parity Generation

This section develops methods for combining the parity generation operations with filter banks, such as shown in Figure 3, forming a cascaded system, depicted in Figure 5. A generic channel with signal value notation overlaid is presented in the middle of this figure. The output of the tth filter, $H_t(Z)$, is denoted by $Y_t(Z) \leftrightarrow \{y_t(r)\}$. The parity output $\{p_t(a)\}$, after downsampling by factor K, may be written in terms of the tth channel signal $\{y_t(r)\}$, which in turn can be expressed using segments of the filter's impulse response.

$$p_t(a) = \sum_{v=0}^{N-1} \sum_{u=-\infty}^{+\infty} x(uN + v) \; g_t^{(v)}(aK - u) \quad (6a)$$

The composite weighting functions $\{g_t^{(v)}(r)\}$ contain every N^{th} sample of the filter weighting, properly offset by index v.

$$g_t^{(v)}(s) = \sum_{r=-\infty}^{+\infty} q(r) h_t((s-r)N - v), \quad (6b)$$

where $v = 0, 1, ..., N-1$. The output sample index a is scaled by K in the argument of $g_t^{(v)}(\;)$ inside the definition of $p_t(a)$, equation (6a), while it is further scaled by N in this definition, equation (6b). The net effect has the input data weighted by values every N^{th} point, in steps of KN with respect to the data indices. The Z transform of the composite impulse responses, equation (6b), will be denoted by $G_t^{(v)}(Z)$.

The real savings in computing the respective channel parities occur for the case of uniform filters at the critically sampled rate, $L = N$. With the filter bank as in Figure 2,

163

the outputs of each $H_t(Z)$ are scaled by a complex phasor as in equation (2). This translates the parity channel output $p_t(a)$ into a modified equations (6).

$$p_t(a) = \sum_{v=0}^{N-1} \sum_{u=-\infty}^{+\infty} [x(uN+v)\, e^{j\frac{f_s}{N}tv}]\, g^{(v)}(ak-u) \quad (7)$$

The uniform filter weighting function $g^{(v)}(s)$ is defined similarly to equation (6b). The complex roots of unity are functions only of the outer index v, and, when all N channels are considered, the complete set of parity values may be calculated by a DFT operation, as described earlier with regard to the polyphase multirate filter banks. The calculation rate is reduced by a factor KN, even though the individual composite channels accept data at intervals of N.

Protecting a Polyphase Filter Demultiplexing System

The parity values are calculated in two ways, one by a parallel composite parity generation process as described in the last section. The second comparable parity values are computed directly from the channel's demultiplexed output. The first set of parities are calculated according to equations (6) employing the composite weighting. The other parity estimates are computed directly from the individual channel outputs using the parity weighting Q(Z). These two versions of $p_t(a)$, labeled $p_t{'}(a)$ and $p_t{''}(a)$ are compared in a totally self-checking comparator. The combined protection system is detailed for generic channels r in Figure 6; identical calculations for each of the N outputs would be made.

The full details of this generalized version of a totally self-checking equality checker [7] are contained in a book chapter [25]. The threshold value Δ in this comparator is selected to allow small differences between the two versions of comparable parity samples, accounting for roundoff noise discrepancies arising because they are computed by different subsystems. The parity weight filters, $G^{(v)}(Z)$ blocks in Figure 6, combine the effects of Q(Z) and H(Z). However, the computational rate is reduced further by a factor of K, making this scheme an efficient protection approach. Since each channel compares a pair of parity values every K^{th} output value, errors are detected with a latency of at most K output samples. The detecting capability of the code is sometimes specified in terms of the minimum distance for a constraint length.

Conclusions and Future Work

This paper has demonstrated how real convolutional codes can be employed efficiently for protecting demultiplexer filter banks. Each demultiplexer output channel has two forms of low rate parity calculation associated with it. One value is computed directly from the output using an FIR parity filter dictated by the structure of a real convolutional code. The memory in the parity filter

is determined by the constraint length of the code while its very favorable downsampled processing rate is governed by the code rate. The other parity value is computed in parallel with the normal processing by a composite filter operating at a reduced rate as governed by the convolutional code choice. These parallel parity calculations can be implemented very efficiently by a polyphase multirate filter bank, virtually identical with the main demultiplexer bank, except operating at a much lower rate.

Research is continuing on the fault tolerance aspects of other subsystems shown in Figure 1. The control sections inherent in each subassembly are not always visible to the system designer. The data level protection techniques promoted in this paper provide coverage for many control action failures. However, there are control steps that are not directly covered, for example, those actions associated with parity comparison results or reconfiguration decisions. There are evolving techniques that can be applied at the microcode level employing embedded checks recomputed by a small hardware monitor at run time and compared [26-27].

The demodulators are very difficult to protect because of their internal phase and timing tracking loops. On the other hand, the forward error correctors (FEC) directly following contain redundancy checks. These devices generally use convolutional codes and implement the Viterbi Algorithm [21-22]. Hardware failures in a channel demodulator appear as channel errors to the respective FEC devices. Hence, if an individual FEC subassembly is fault-free, any errors detected by it lead to suspecting failures in the preceding DEMOD unit. This is a "sandwich" approach to fault tolerance. If succeeding and preceding units are fault-free and errors are detected by the protected succeeding unit, the item in the middle of the "sandwich" suspect. Work is progressing on introducing data level fault tolerance in Viterbi algorithm type decoders, primarily by attaching real parities to groups of path metrics [21-22]. Again, the principle of algorithm-based fault tolerance is used. Real parity values are computed and recomputed and then compared.

References

[1] M. Bellanger, *Digital Processing of Signals: Theory and Practice* (2nd Edition). New York: John Wiley & Sons, 1989.

[2] M. Bellanger and J. L. Daguet, "TDM-FDM Transmultiplexer: Digital Polyphase and FFT," *IEEE Transactions on Communications*, Vol. COM-22, pp. 1199-1205, 1974.

[3] M. Bellanger, G. Bonnert and M. Coudreuse, "Digital Filtering by Polyphase Network: Application to Sample-Rate Alteration and Filter Banks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-24, pp. 109-114, 1976.

[4] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs: Prentice-Hall, 1983.

[5] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs: Prentice-Hall, 1989.

[6] D. K. Pradhan, Editor, *Fault-Tolerant Computing Theory and Techniques*, Vol. 1., Englewood Cliffs: Prentice-Hall, 1986.

[7] J. Wakerly, *Error Detecting Codes, Self-Checking Circuits and Applications*. New York: North-Holland, 1978.

[8] B. W. Johnson, *Design and Analysis of Fault-Tolerant Digital Systems*. Reading, MA: Addison-Wesley Publishing Company, 1989.

[9] K.-H. Huang and J. A. Abraham, "Algorithm-Based Fault Tolerance for Matrix Operations," *IEEE Transactions on Computers*, Vol. C-33, pp. 518-528, 1984.

[10] J.-Y. Jou and J. A. Abraham, "Fault-Tolerant Matrix Arithmetic and Signal Processing on Highly Concurrent Computing Structures," *Proceedings of the IEEE (Special Issue on Fault Tolerance in VLSI)*, Vol. 74, pp. 732-741, 1986.

[11] J. A. Abraham, "Fault Tolerance Techniques for Highly Parallel Signal Processing Architectures," *SPIE Highly Parallel Signal Processing Architectures*, Vol. 614, pp. 49-65, 1986 (K. Bromley, Editor).

[12] C. J. Anfinson and F. T. Luk, "A Linear Algebraic Model of Algorithm-Based Fault Tolerance," *IEEE Transactions on Computers*, Vol. C-37, pp. 1599-1604, 1988.

[13] F. T. Luk and H. Park, "An Analysis of Algorithm-Based Fault Tolerance Techniques," *SPIE Advanced Algorithms and Architectures for Signal Processing*, Vol. 696, pp. 222-227, 1986.

[14] W. G. Bliss, "Area-Time Efficient and Fault-Tolerant VLSI Arrays for Digital Processing," Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Colorado, 1988.

[15] C. J. Anfinson, R. P. Brent and F. T. Luk, "A Theoretical Foundation for the Weighted Checksum Scheme," *SPIE Advanced Algorithms and Architectures for Signal Processing*, Vol. 975, pp. 10-18, 1988.

[16] R. P. Brent, F. T. Luk and C. J. Anfinson, "Choosing Small Weights for Multiple Error Detection," SPIE High Speed Computing, Vol. 1058, pp. 16-1–16-7, 1989.

[17] F. T. Luk, "Algorithm-Based Fault Tolerance for Parallel Matrix Equation Solvers," *SPIE Real-Time Signal Processing*, Vol. 564, pp. 49-53, 1985 (W. J. Miceli and K. Bromley, Editors).

[18] T. G. Marshall, Jr., "Coding of Real-Number Sequences for Error Correction: A Digital Signal Processing Problem," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-2, pp. 381-392, 1984.

[19] J. K. Wolf, "Redundancy, the Discrete Fourier Transform, and Impulse Noise Cancellation," *IEEE Transactions on Communications*, Vol. COM-31, pp. 458-461, 1983.

[20] T. G. Marshall, Jr., "Real Number Transform and Convolutional Codes," *Proceedings 24th Midwest Symposium on Circuits and Systems*, Albuquerque, NM, pp. 650-653, June 1981.

[21] W. W. Peterson and E. J. Weldon, Jr., *Error-Correcting Codes* (Second Edition). Cambridge, MA: The MIT Press, 1972.

[22] S. Lin and D. J. Costello, Jr., *Error Control Coding Fundamentals and Applications*. Englewood Cliffs: Prentice-Hall, 1983.

[23] V. S. S. Nair and J. A. Abraham, "Real Number Codes for Fault-Tolerant Matrix Operations on Processor Arrays", *IEEE Transactions on Computers*, Vol. C-39, pp. 426-435, 1990.

[24] J. Hagenauer, "High Rate Convolutional Codes with Good Distance Profiles," *IEEE Transactions on Information Theory*, Vol. IT-23, pp. 615-618, 1977.

[25] G. R. Redinbo "Real Codes in the Fourier Domain for Fault-Tolerant Signal Processing." Chapter in *Spectral Techniques: Theory and Applications*. North-Holland: Amsterdam, 1991. (Moraga and Creutzburg, Editors.)

[26] Kent D. Wilken, "Optimal Signature Placement for Processor-Error Detection using Signature Monitioring", *Proceedings Twenty-First International Symposium on Fault-tolerant Computing*, Monteral, Canada, pp.180-187, June 1991.

[27] K. Wilken and J. Shen, "Continuous Signature Monitoring: Low-cost Concurrent-Detection of Processor Control Errors", *IEEE Transactions on Computer-Aided Design*, col. CAD-9, pp.629-641, June 1990.
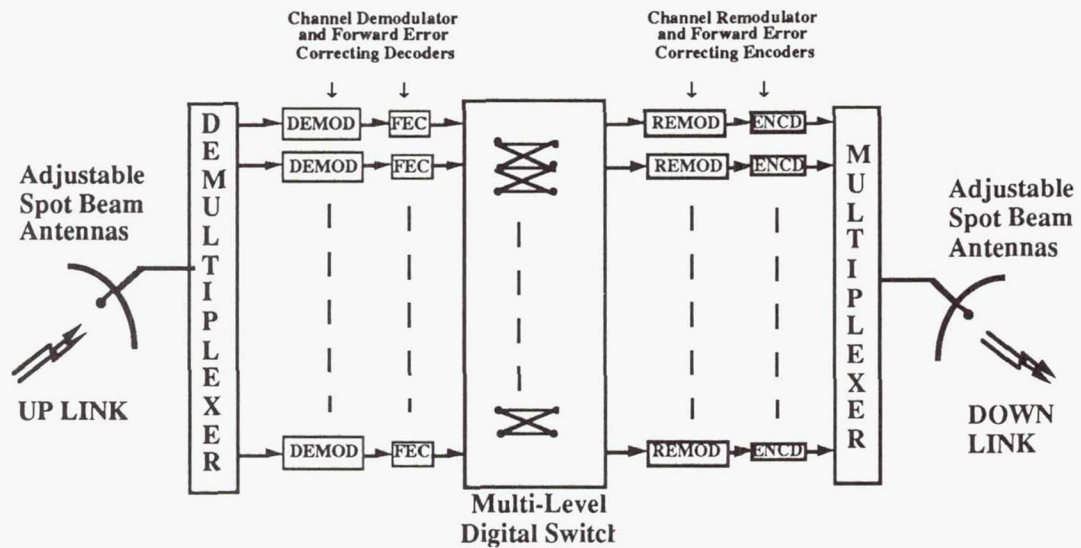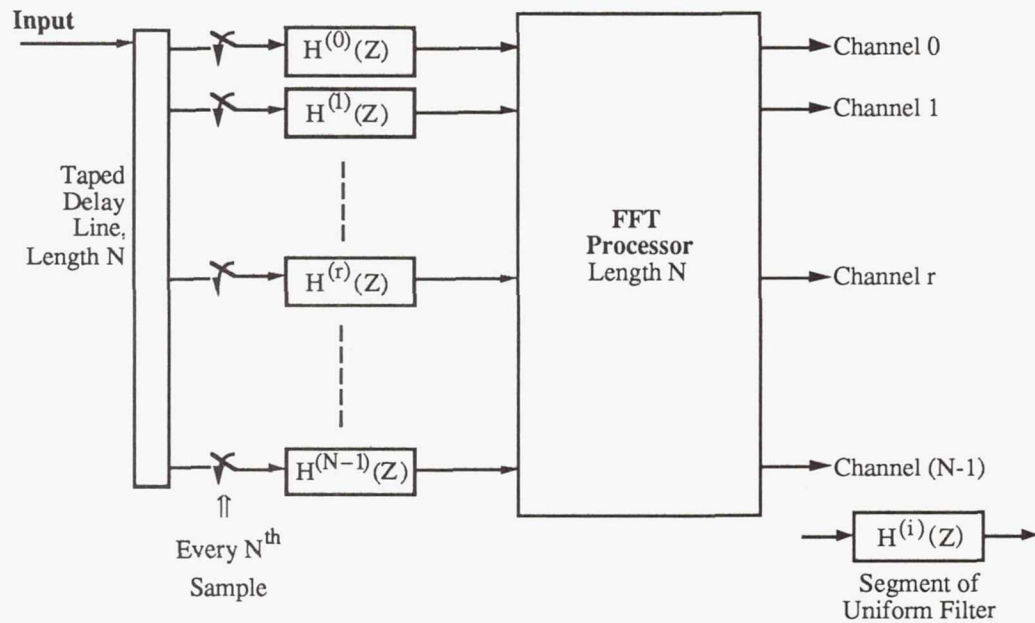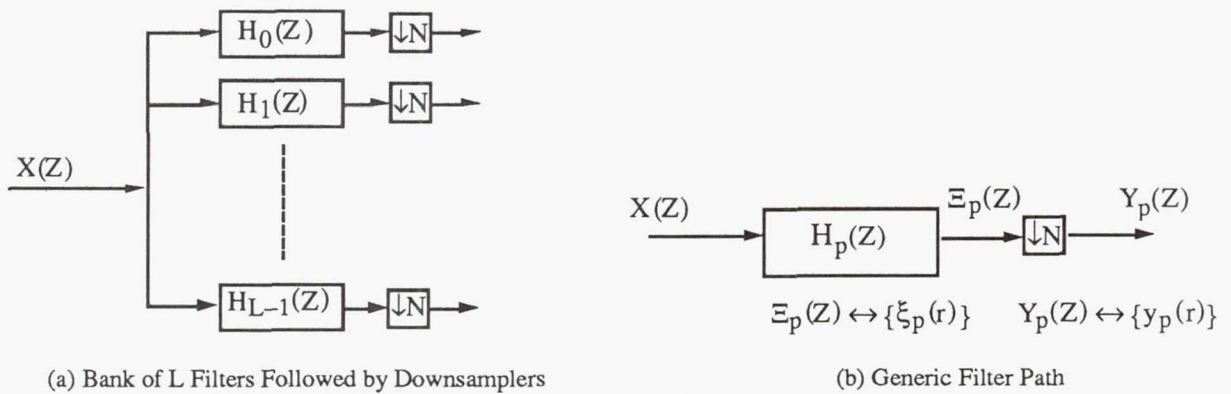
Figure 1: Typical Communication Satellite Subsystems

Figure 2: Polyphase Multirate Filter Bank

(a) Bank of L Filters Followed by Downsamplers   (b) Generic Filter Path
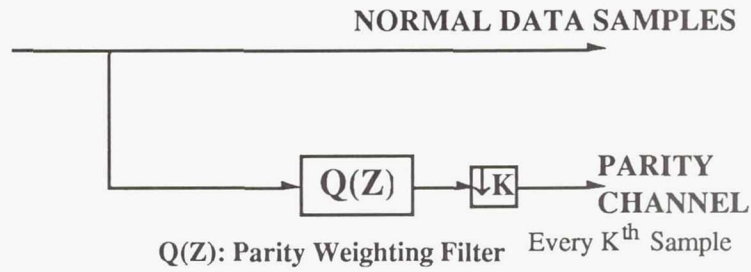
Figure 3: General Analysis Bank

NORMAL DATA SAMPLES

Q(Z): **Parity Weighting Filter** Every K^{th} Sample

Figure 4: Parity Generation in a Rate (K/K+1) Systematic Convolutional Code
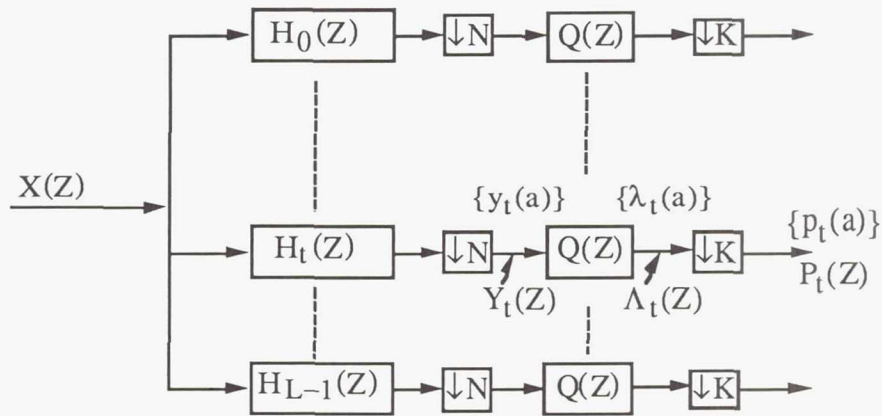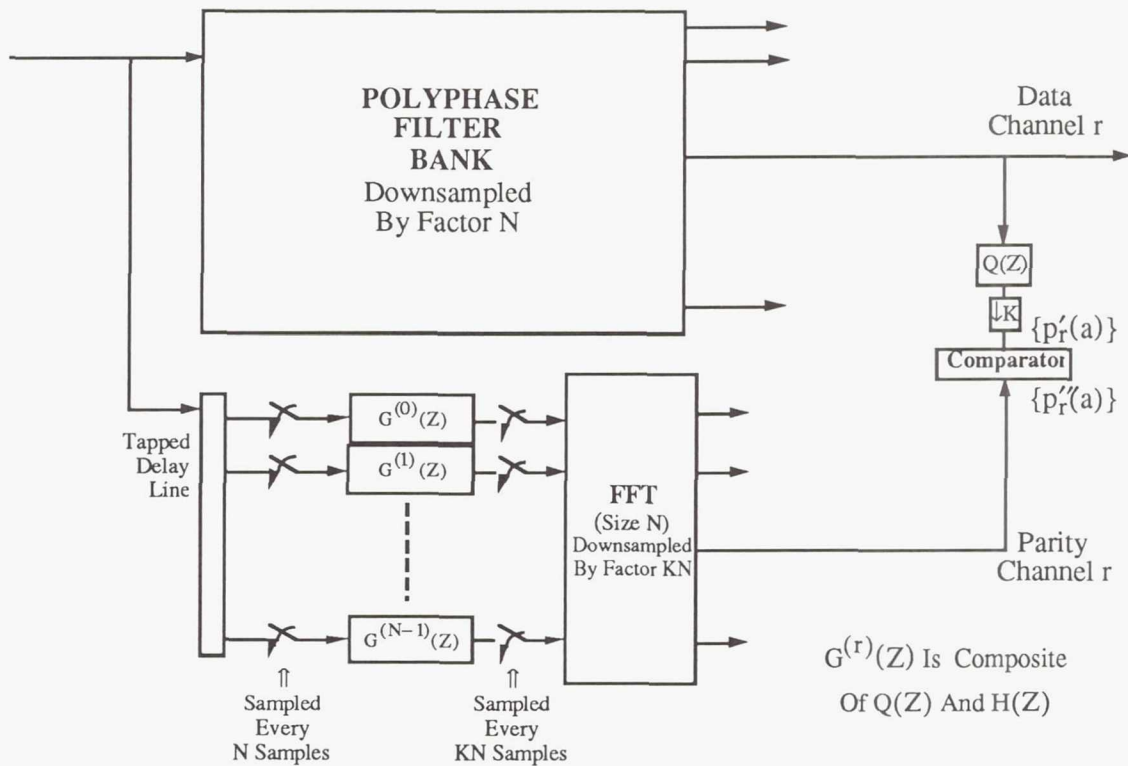

Figure 5: Parity Generation for Analysis Bank Outputs

Figure 6: Protection of Demultiplexer Channels

167

Page intentionally left blank

ON-BOARD DEMUX/DEMOD

S. Sayegh, M. Kappes, J. Thomas, J. Snyder,
M. Eng, J. Poklemba, M. Steber, and G. House
COMSAT Laboratories
Clarksburg, Maryland 20871

Abstract

To make satellite channels cost competitive with optical cables, the use of small, inexpensive earth stations with reduced antenna size and high powered amplifier (HPA) power will be needed. This will necessitate the use of high e.i.r.p. and gain-to-noise temperature ratio (G/T) multibeam satellites. For a multibeam satellite, on-board switching is required in order to maintain the needed connectivity between beams. This switching function can be realized by either an receive frequency (RF) or a baseband unit. The baseband switching approach has the additional advantage of decoupling the up-link and down-link, thus enabling rate and format conversion as well as improving the link performance. A baseband switching satellite requires the demultiplexing and demodulation of the up-link carriers before they can be switched to their assigned down-link beams. This paper discusses principles of operation, design and implementation issues of such an on-board demultiplexer/demodulator (bulk demodulator) that was recently built at COMSAT Laboratories.

1. INTRODUCTION

A multiyear effort was undertaken at COMSAT Laboratories to investigate the on-board demultiplexer/demodulator concept to determine its feasibility, identify critical technologies, and assess the potential of developing these technologies to a level capable of supporting a practical, cost-effective on-board implementation. An important part of the effort was a review of the advances that can be expected to occur in the critical digital component areas in terms of power, mass, size, speed, and radiation resistivity of the digital, logic, and memory components from which the processor is to be fabricated.

A baseline system of the demultiplexer/demodulator was defined and its performance evaluated by analysis and computer simulations. A digital implementation was selected to provide the flexibility that permits the on-board processor to accommodate different types of multichannel frequency-division multiple access (FDMA) carriers simply by changing its computational rules and organization. This permits the rules and organization of each processor to be modified to accommodate variations in the number and bandwidths of carriers over the lifetime of the satellite or to accommodate different applications of the same type of satellite.

A block diagram of the overall system and test setup is shown in Figure 1. The system uses the frequency-domain filtering approach to demultiplexing and a shared high-speed coherent demodulator. The fast Fourier transform (FFT)-based demultiplexer is capable of processing a large number of carrier types and bit rates. The demultiplexer output is fed into an interpolating filter whose task is to deliver 2 samples per symbol to a shared variable bit rate digital demodulator that operates on a number of different carriers in a round-robin fashion. The COMSAT digital processor performs demultiplexing/demodulation and associated filtering and control for a number of carriers occupying a bandwidth of 20 MHz. The architecture used in this system is very flexible, allowing in-orbit frequency plan reconfiguration under ground command.

Most of the hardware has been implemented in low-power complimentary metal-oxide semiconductor (CMOS) circuitry. Several other important developments contributed to very substantial reductions in the power, mass, and size of the processor. An application-specific integrated circuit (ASIC) gate array chip that performs the interstage reordering in the FFT pipeline was designed and developed. This contributed to better than an order of magnitude reduction in power and mass as compared with a discrete large-scale integration (LSI) implementation. A method for sharing a single pipeline inverse FFT processor among the different carriers was conceived. By interleaving the frequency samples of those carriers at the input to the inverse FFT (IFFT) processor and selectively bypassing butterfly operations, carriers of different bandwidths can be handled simultaneously in the shared pipeline. This obviates the need for a separate IFFT processor for each carrier. A novel PROM-based approach was implemented for the acquisition section of the shared digital demodulator, significantly reducing the required hardware.

The demultiplexer/demodulator presented above has been constructed and tested at COMSAT Laboratories and is now operational. System performance evaluation in terms of bit error rate measurements are presented in this paper.

2. FREQUENCY DOMAIN FILTERING

An FFT/IFFT frequency-domain filtering architecture was selected for the demultiplexer. FFT/IFFT frequency-domain filtering method basically consists of convolving the composite frequency multiplexed signal with a bank of filters using an overlap-and-save technique. It computes the desired linear convolutions in terms of circular convolutions. The circular convolutions are computed by transforming the time-domain quantities to be convolved to the frequency domain, multiplying the resulting frequency coefficients across the overall spectrum by any desired filter functions and transforming back to the time domain. Specifically, to obtain carrier k, the frequency multiplexed signal is transformed to the frequency domain by an FFT, multiplied by the frequency response of filter k (typically a square-root Nyquist that serves the double purpose of demultiplexing and matched filtering), and the product is then

transformed back via an IFFT to recover the time-domain waveform. Therefore, to obtain the individual baseband signals, the number of inverse transforms to be performed equals the number of carriers N. To minimize the amount of computation involved, the IFFT performed on a given carrier should only cover the frequency band occupied by that carrier. Thus, different carrier bandwidths will result in different IFFT sizes.

Figure 2 summarizes the frequency-domain demultiplexing approach. An important consideration here is the size of the Fourier transform that has to be performed. If the filter impulse response has L coefficients and 50-percent overlap is used between blocks, then the size of the Fourier transform to be performed is $2L$. Note that as the overlap between blocks increases, so does the number of computations per output sample. On the other hand, if the overlap decreases, then the size of the Fourier transform, and hence the memory size, increases. A 50-percent overlap achieves an almost optimum tradeoff between computational and memory requirements and is very simple to implement.

A pipeline FFT processor is an efficient way of performing the needed high-speed, real-time Fourier transforms by distributing the processing among several computational elements. It has a compact and modular structure and is well suited for very large-scale integration (VLSI) implementation. The pipeline processor consists of two building blocks: butterfly computational elements and delay-switch-delays (DSDs). The computational elements perform the necessary butterfly computations of the Cooley Tukey FFT algorithm. The DSDs consist of shift registers first-in first-out (FIFOs) and switches. Their function is to present the samples to the butterfly computational elements at the right place at the right time.

A radix-2 or a radix-4 implementation may be used for the FFT/IFFT pipelines. Although the radix-4 butterfly computations are more involved than those of the radix-2 (three complex multiplications vs one for the radix 2), the number of butterfly stages is half that required for a radix 2 and they operate at half the speed. Therefore the number of complex multiplications per second is 25 percent smaller for a radix-4 implementation. The choice of radix for the FFT is thus a tradeoff between speed and additional hardware. If the speed requirement can be satisfied by either implementation, then radix 2 may be the preferred choice. At the time the proof-of-concept (POC) model was designed, the radix-4 pipeline, shown in Figure 3, was chosen because of speed considerations.

As more highly-integrated devices become available, however, this choice must be reconsidered. A radix-4 butterfly chip operating at one speed and a radix-2 butterfly chip operating at twice the speed can each handle the same data rate. Thus, the answer to which is best turns to such factors as package size and power consumption, with consideration of the fact that the radix-2 pipeline requires twice as many butterflies.

As signal processing chips advance beyond the basic butterfly operation, the additional functions they include and the means of controlling them must be considered to determine whether competing devices are more or less desirable. For example, one manufacturer may offer on-chip coefficient memory while another may not. A third may have coefficient memory that requires more off-chip control to utilize for our application. Therefore, the best architecture depends upon the total board area and power consumption required to perform a complete function (such as an FFT) at a particular speed.

The FFT pipeline in the POC model is capable of accepting four complex input samples and providing four complex output samples during each 11.52-MHz clock period. Thus, a 256-point FFT is computed every 64 clock periods or 5.6 μsec. By extending the length of the pipeline, a 1024-point FFT could be computed every 22 μsec.

A reduction in these times by a factor of approximately two would be a desirable objective for the near future in order to double the maximum IF bandwidth to about 40 MHz (which corresponds to a pipeline data rate of 80 x 10^6 complex samples/sec.). The long-term goal is an additional factor-of-two improvement to permit direct processing of IF bandwidths as great as 80 MHz.

The DSD is one of two key elements used to implement the pipeline FFT and IFFT processors (the other being the butterfly arithmetic processor). Due to the complexity (large amount of hardware) of the circuit, it is more practical to implement it with ASIC technology. COMSAT Laboratories has developed this DSD ASIC chip as part of the demultiplexer/demodulator program. A detailed description of the COMSAT developed DSD is now presented.

To implement one complete DSD function, eight ASIC chips and a small amount of discrete logic integrated circuits (ICs) are needed. In the FFT processor, one complete DSD is used between two butterfly elements and its function is to reorder the samples in its input data streams appropriately for the butterfly that follows it. The reordering process is achieved by using two sets of delay elements and one set of switch elements. The first set of delay elements are used to shift the input streams appropriately in time, then the switch elements interchange the samples in a predetermined fashion, and finally the second set of delay elements are used to shift the samples back in time appropriately. The DSD ASIC is hardware-programmable for the particular stage of the FFT or IFFT processor in which it is used. Specifically, there are four possible configurations for the DSD, three of which are for radix-4 transforms and one for the radix-2 case used in the IFFT processor. In the radix-2 case, the DSD treats the four input streams as two groups of two input streams. The DSD configured for the stage closest to the output of the processors has the smallest amount of delays and the highest switching rate.

The functional block diagram of the DSD ASIC is shown in Figure 4. The number on the right side of the delay element blocks indicate the delay values associated with the particular input data stream. For example, the four inputs X11-X14 always have delay values of 0, the four inputs X21-X24 can have delay value of 1, 4, 8 or 16 clock cycles. For any one configuration selected, only one set of

delay values are used for the input and output data streams. For example, if the DSD is used in the second to the last stage of the FFT processor, the delay values 0, 4, 8 and 12 are used. Specifically, the signals X11-X14 and Y41-Y44 assume the delay value of 0, the signals X21-X24 and Y31-Y34 assume the delay value of 4, the signals X31-X34 and Y21-Y24 assume the delay value of 8, and so on.

The switch elements in the DSD perform the function of routing the incoming signals to the appropriate outputs of the switch elements. With reference to Figure 5, there are two configurations for the switch elements, one of them is the 2-state and the other is the 4-state case, and they are used for the radix-2 and radix-4 applications, respectively. For the 2-state case, in state '0', the inputs to the DSD go straight through it, and in state '1', data streams at ports INA and INB are interchanged and data streams at ports INC and IND are interchanged. For the 4-state case, the situation is slightly more complicated, and the actions taken by switch elements in each of the four states are shown in Figure 5. For any one of the two switch configurations selected, the switch elements always go through the same states. The only difference when the DSD is used in different stage of the FFT or IFFT processors is the rate at which the states are switched. Specifically, the DSD closest to the FFT processor output has the highest switching rate, and it switches state every clock cycle. The DSD in the preceding stage switches states once every four clock cycles, and so on.

With reference to Figure 6 and the functional details of the DSD ASIC mentioned above, the implementation description is presented next. The delay elements are implemented using shift registers and 4-to-1 multiplexers. For a particular data stream, the possible delays of the data is achieved by connecting the outputs of the shift registers from the appropriate output stages to the inputs of the associated 4-to-1 multiplexer, and by selecting one of the four inputs as the output.

The switch elements are also implemented using 4-to-1 multiplexers, and by selecting the appropriate inputs, the data interchange as indicated in Figure 5 can be accomplished. The controller of the DSD ASIC is responsible for providing all the timing and control signals for the shifting and multiplexing operations within the ASIC.

Whereas a single FFT is performed on the composite frequency-division multiplexing (FDM) signal, the IFFTs are performed on an individual carrier basis. The case of mixed carrier sizes (narrowband, and wideband) is readily handled by performing inverse transforms of larger sizes for the wideband carriers. In order to avoid the duplication of hardware, a common pipeline, capable of performing transforms of various sizes under software control without the need for physically adding or removing any modules, is desirable.

When a pipeline is dedicated to performing an FFT of a given size, the number of stages in the pipeline is fixed and the twiddle factors of the butterfly computations, as well as the switching times and delays of the DSDs are readily available. The pipeline can be modified to perform transforms of various sizes simultaneously. The needed modifications are performed dynamically (using a few control signals) to allow the pipeline to constantly alter its function in real-time to accommodate the various transformation sizes required. By properly ordering the input data to the pipeline and bypassing some arithmetic modules for the smaller size transforms, any mixture of IFFTs whose sizes are a power of the pipeline's radix can be performed without requiring any changes to the simple and regular action of the DSD, as illustrated in Figure 7.

3. INTERPOLATION

Because the FDMA signal consists of asynchronous carrier transmissions, the samples at the demultiplexer output must be interpolated before being presented to the demodulator. The interpolating filter module (IFM) which connects the demultiplexer output to the demodulator input performs two functions. First, it adjusts the number of samples per symbol for each carrier from an arbitrary value near two to exactly two. Second, it adjusts the sampling point for each carrier to coincide with the peaks and zero crossings of the signal. It performs both functions by means of adjusting the phase shift of a simple finite impulse response (FIR) digital filter. The control signals for adjusting the number of samples per symbol are generated locally and asynchronously and are added to the accumulated clock error fed back from the demod to produce a composite control signal proportional to the instantaneous phase adjustment for the current sample. This signal is fed to the FIR filter.

Figure 8 shows the shared control circuitry of the IFM. The upper part of the diagram shows the circuitry required to generate the coefficient programmable read-only memory (PROM) addresses as well as general control signals. Each carrier address counter keeps track of the location within the current phase plan used for correcting the number of samples per symbol. This address is fed to a phase plan lookup PROM for each type of carrier in use. These PROMs are shared by different carriers of the same bit rate and coding scheme. This signal is then added to the output of the clock error accumulator for each channel and applied to the coefficient lookup PROMs to obtain the coefficients for the FIR filter. As a practical matter, the coefficient PROMs shown are duplicated on the second board to avoid too many board-to-board connections.

Figure 9 shows the nonshared circuitry of the IFM, i.e., circuitry that must be repeated for the I and Q channel. The shift register, multipliers, and adders constitute the basic FIR filter circuit. The data buffer and related control circuitry provide samples on demand to the input of the FIR filter. This is necessary in cases where the samples in the shift register must be reused to generate two output samples. In other cases where the current contents of the shift register are not required for an output sample the outputs are simply marked as invalid. This peculiarity occurs due to the fact that the number of input samples does not match the number of output samples.

In its current configuration, the entire IFM occupies two 9Ux440 wirewrap cards and uses predominately high-speed

CMOS digital logic devices including the LSI Logic L64012 multiplier IC and the Logic L4C381 accumulator IC. The first board contains the shared control circuitry and the I channel filter while the second board contains the Q channel filter.

4. DEMODULATOR

The on-board demodulator operates on a multiple set of quadrature phase shift keying (QPSK) modulated, asynchronous carriers in a TDM format, where the incoming TDM data packets are typically a fraction of the transmitted burst. In this manner, the demodulator processes only a few symbols for a given carrier, stores the results, and preloads its registers with the appropriate sample values for the upcoming carrier. The sample rate entered into the demodulator for all carriers is two samples/symbol. Recall that the sample frequency for all carriers is the same as their symbol rates after they have been warped in the FDM/TDM conversion. Symbol timing feedback from the tracking loop to the preceding interpolating filter places the two samples into the demodulator at the data-detection and symbol-transition points. The receive Nyquist data shaping has already been done in the receive filter module. However, the sample values at the data-detection points are modulated by a beat note between the actual incoming center frequency and the front-end down-conversion local oscillator. A carrier-phase rotator, which is effectively a 2 x 2 matrix multiplication of the beat modulated I and Q channels, is employed to remove the beat as follows:

$$
\begin{bmatrix} I_k \\ Q_k \end{bmatrix} = \begin{bmatrix} \cos(\widehat{\theta}) & -\sin(\widehat{\theta}) \\ \sin(\widehat{\theta}) & \cos(\widehat{\theta}) \end{bmatrix} \begin{bmatrix} I_k' \\ Q_k' \end{bmatrix}
$$

where $\widehat{\theta}$ is the carrier tracking loop output phase estimate.

With a "0101" acquisition preamble in both channels there is a potential 180° ambiguity in the recovered carrier phase, which is resolved by means of the unique word (UW). The UW pattern is the same in both channels, so binary decisions used to increase detection reliability.

There are two phase-locked tracking loops in the demodulator for carrier phase and symbol timing. The carrier-phase tracking is second order to account for frequency offsets, whereas the symbol timing is first order and only tracks slow-varying phases. Multiplier-accumulators (MACs) are used to implement the digital tracking loops. The accumulators are preloaded with initial-phase or frequency information, whichever is appropriate, from the acquisition estimator circuitry. In this manner, the phase-locked tracking loop synchronization in burst mode can be expedited and more reliable. In terms of the second order loop parameters, the phase and frequency multiplier gains for carrier tracking are selected respectively as

$$
K_\theta = 2\zeta W_n T_s
$$

$$
K_{\Delta\theta} = (W_n T_s)^2
$$

where ζ is the damping ratio, W_n is the natural frequency, and T_s is one symbol time interval. For the first order symbol timing loop the multiplier gain is

$$
K_\tau = (W_n T_s)
$$

Initial estimates for carrier phase and frequency as well as symbol timing are computed in the acquisition estimate processor as shown in Figure 10, and briefly described as follows. Incoming I and Q channel samples are multiplied by a bipolar alternating sequence to remove the preamble modulation and averaged to improve their signal-to-noise (S/N) ratio. This yields four outputs, namely, even and odd sums in both the I and Q channels. The sums are taken twice, over the first and second halves of the preamble. The carrier-phase error may be found from

$$
\widehat{\theta} = \tan^{-1}\left(\sqrt{\frac{Q_e^2 + Q_o^2}{I_e^2 + I_o^2}} \right) - 45° \, \text{sgn} \, (I_e Q_e + I_o Q_o)
$$

Similarly, the symbol timing error can be related as

$$
\widehat{\tau} = \frac{2}{\pi R_s} \left| \tan^{-1}\left(\sqrt{\frac{I_o^2 + Q_o^2}{I_e^2 + Q_e^2}} \right) - 45°\left[1 - \text{sgn} \, (I_e I_o + Q_e Q_o) \right] \right|
$$

In both cases, the primary estimate can be found from a lookup table of the inverse tangent of a ratio of squares, and the phase ambiguity can be determined from looking up the sign of a sum of products. Since these computations are only required at a rate of twice per preamble, common processing elements can be used where the differences between phase and timing are incorporated into the final value lookup tables. The final value of the carrier phase at the end of the preamble is

$$
\widehat{\theta}_0 = \widehat{\theta}_2 + \left(\frac{\widehat{\theta}_2 - \widehat{\theta}_1}{P/2} \right) \cdot (P/4), \text{ modulo } 180°
$$

where P/2 is half the preamble length.

The carrier frequency offset estimate is determined as

$$
\widehat{w}_o = \frac{\widehat{\theta}_2 - \widehat{\theta}_1}{P/2}
$$

Lastly, the timing estimates are averaged as

$$
\widehat{\tau}_0 = (\widehat{\tau}_1 + \widehat{\tau}_2)/2
$$

5. PERFORMANCE

The BER performance of the on-board demultiplexer/demodulator processor has been measured using the setup shown in Figure 1. Four modulators are used on the transmit side to generate FDMA/TDMA test signals. All four of the modulators are capable of variable bit rate operation and have synthesized carriers so that a wide variety of frequency plans can be generated. The fourth modulator can

be used as an interfering burst for TDMA measurements. Noise is added at the 140-MHz IF to the combined modulator signals before processing by the demux/demod. The BER of any one of the demodulated channels is measured by the performance monitor by comparing the incoming data with a stored version of the transmitted data. Synchronization is provided by the UW detect signal from the demodulator for the selected channel.

To evaluate the performance of the on-board processor carriers corresponding to 1.544 Mbit/s with rate 3/4 and 1/2 coding and 2.048 Mbit/s with 3/4 coding were utilized. As a baseline, the carriers were first processed individually providing single carrier performance. Next, all three carriers were generated and supplied to the processor, but the IFM was set up to process only one of the carriers. This selection effectively separates the demultiplexing and demodulation functions of the processor so that implementation degradations can be isolated to individual subsystems. Finally, all three signals were allowed to pass through the entire system with the BER monitor selecting one of the three signals. A summary of the performance for 1.544 Mbit/s carrier with rate 3/4 coding for the three setups is shown in Figure 11. As can be seen from this figure, there is a small amount of degradation when the three carriers are introduced relative to the single carrier performance, but very little additional degradation when all of the signals are being processed by the IFM and demodulator. This degradation is thought to be due to a slight nonlinear operation of the demultiplexer front-end and is being investigated. In addition, some flaring of the data occurs at the lower error rates for all of the curves resulting from low-level interference effects. Overall,

the BER performance data provides validation of both the overall demultiplexer/demodulator structure and the selections of bit resolutions made early in the program.

6. CONCLUSIONS AND SUMMARY

An architecture for implementing an on-board flexible demultiplexer/demodulator was presented. The architecture is based on a frequency domain filtering approach to demultiplexing an up-link FDMA signal consisting of a mixture of carriers of different bit rates was presented. Specially designed FFT pipeline processors were used for this purpose. An ASIC chip designed at COMSAT Laboratories as a critical part of the FFT/IFFT processor was described. A digital demodulator architecture that operates on the interpolated demultiplexer output was presented. A survey of current technology illustrated that for the near future high-speed low-power digital signal processing will be mainly based on Si technologies (CMOS and CMOS/silicon-on-sapphire [SOS]). Based on COMSAT's experience with POC developments of processors similar to the ones discussed in this paper, as well as projections of technology, it is estimated that an 80-MHz fully-digital, very-flexible flyable processor is an achievable goal for the late 1990s. Such a processor is projected to consume only 25 W and have a mass under 5 lb.

7. ACKNOWLEDGMENTS
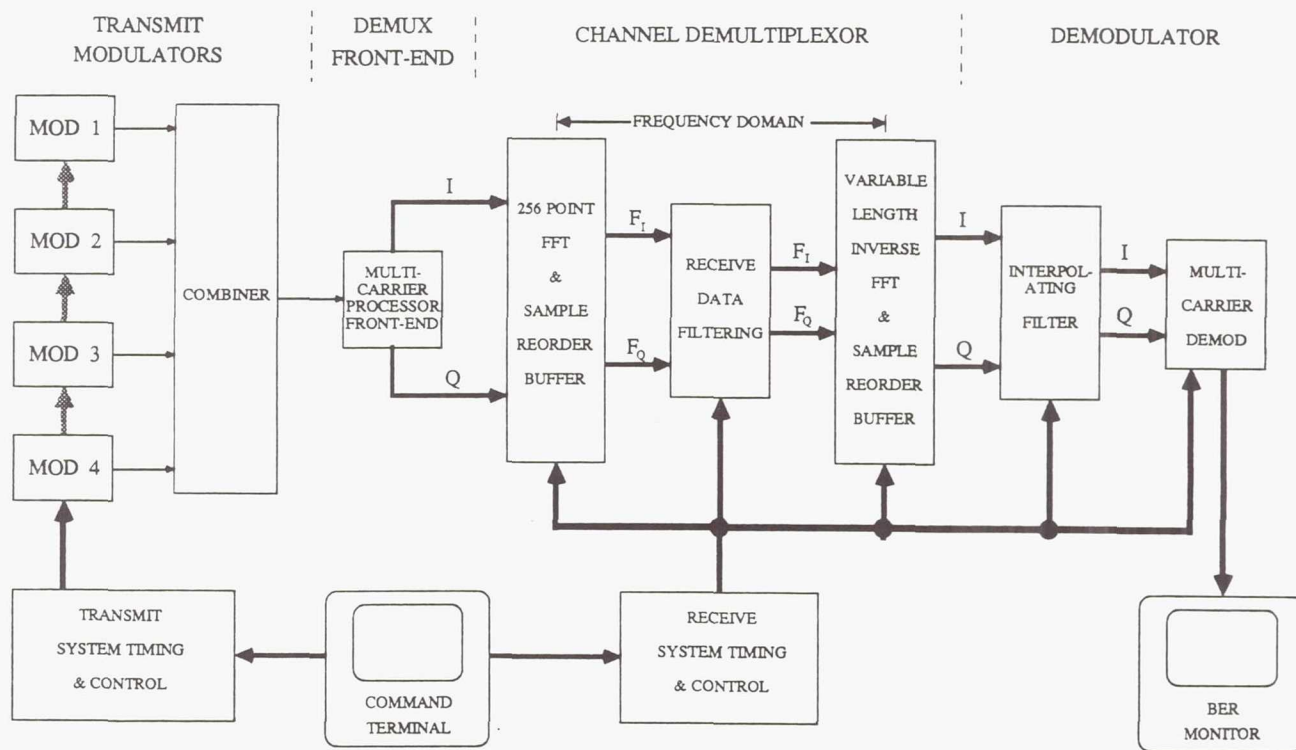
Figure 1. System Block Diagram

173
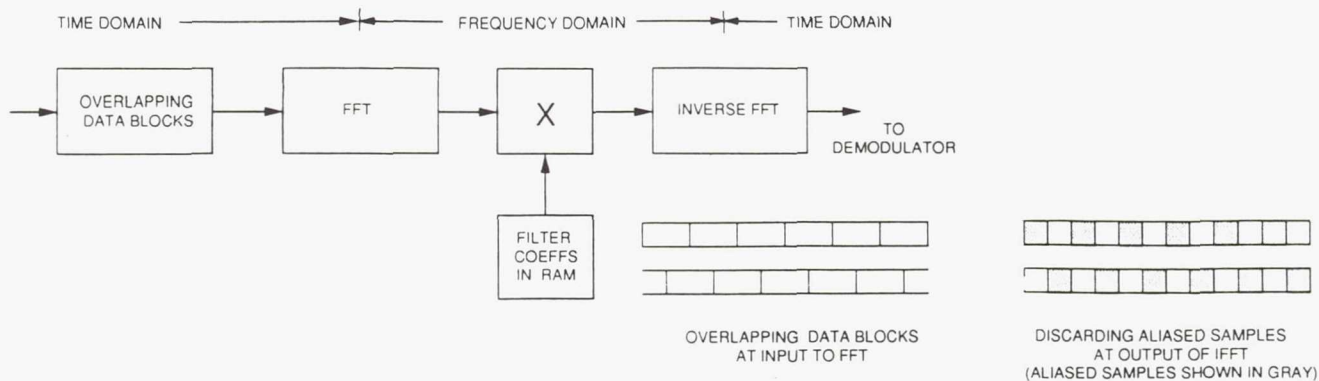
Figure 2. Frequency-Domain Filtering Approach

BF : BUTTERFLY
DSD : DELAY/SWITCH/DELAY
CM : COEFFICIENT MEMORY
kD : k SAMPLES DELAY
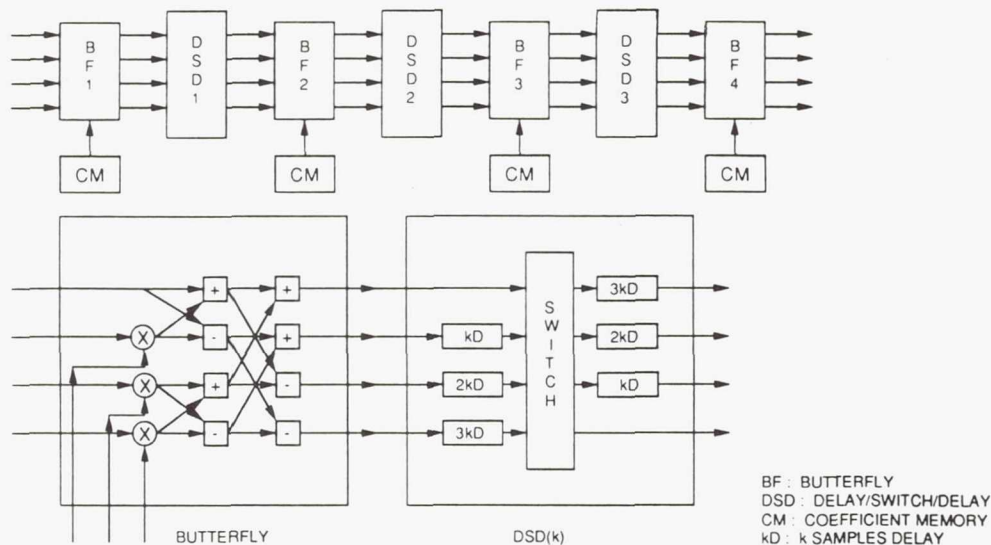
Figure 3. Four-Stage Radix 4 FFT Pipeline
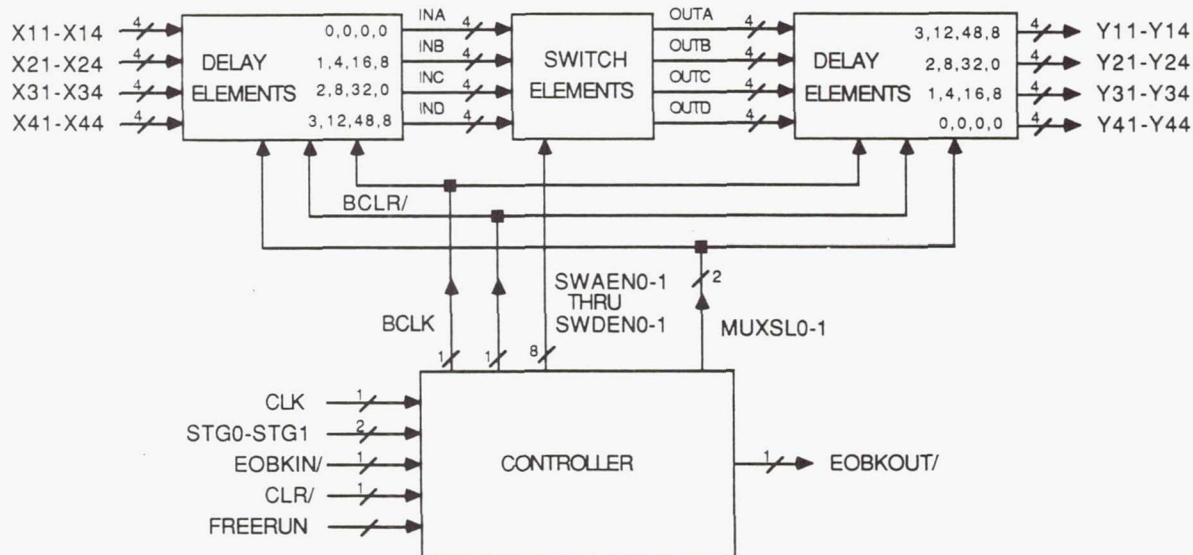
Figure 4. 4-Bit Wide Delay-Switch-Delay ASIC Functional Block Diagram

174

Figure 5. 4-Bit Wide Delay-Switch-Delay ASIC Switch State Diagram
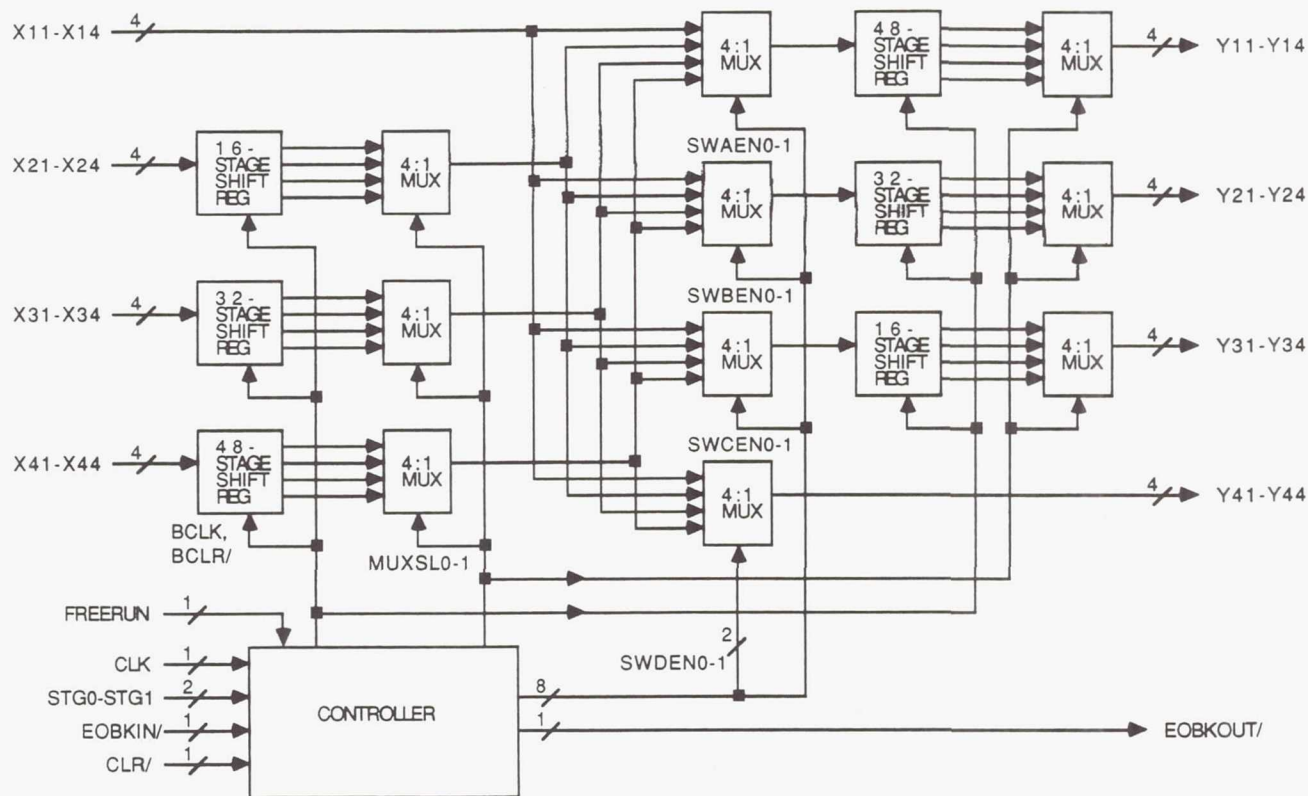
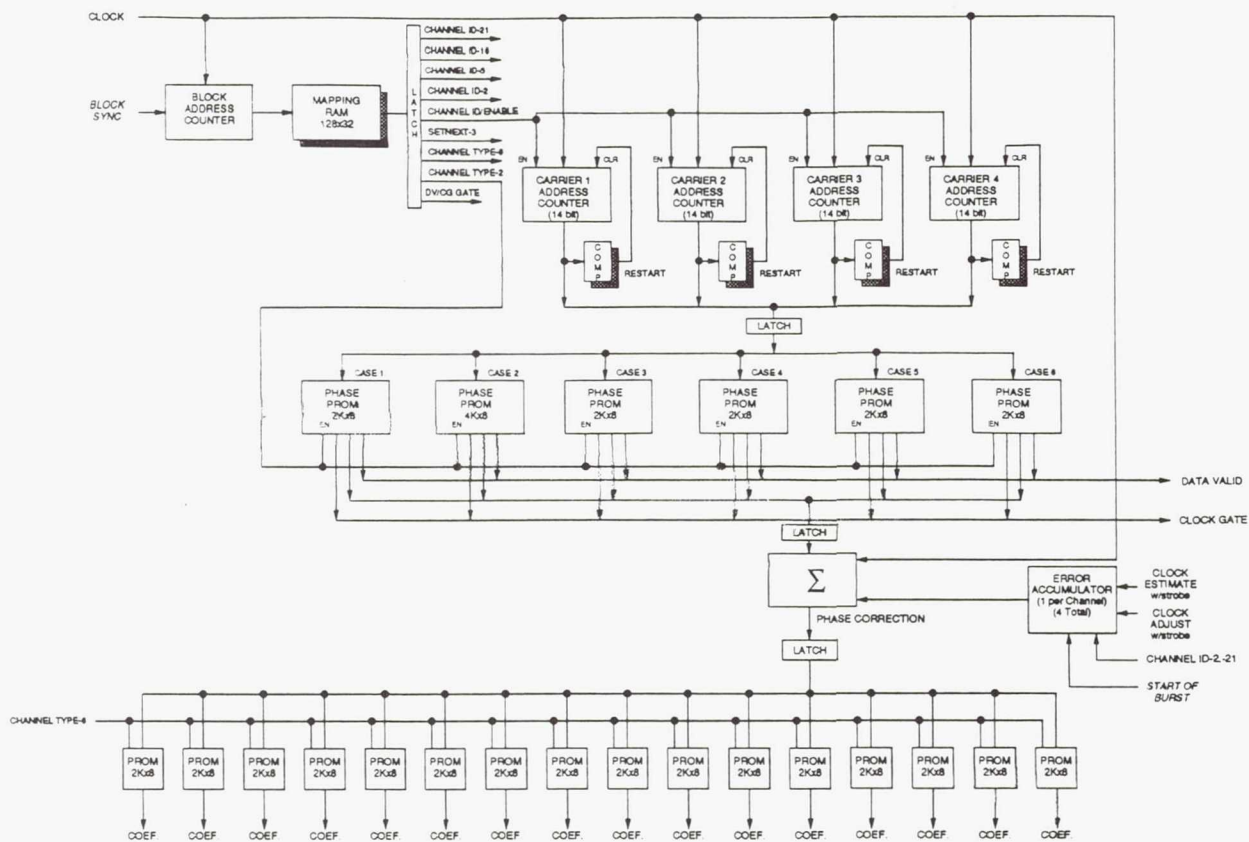Figure 6. 4-Bit Wide Delay-Switch-Delay ASIC Implementation Block Diagram

KEY:

32-point transform

16-point transforms

8-point transforms

4-point transforms

2-point transforms

X   Bypassing butterfly
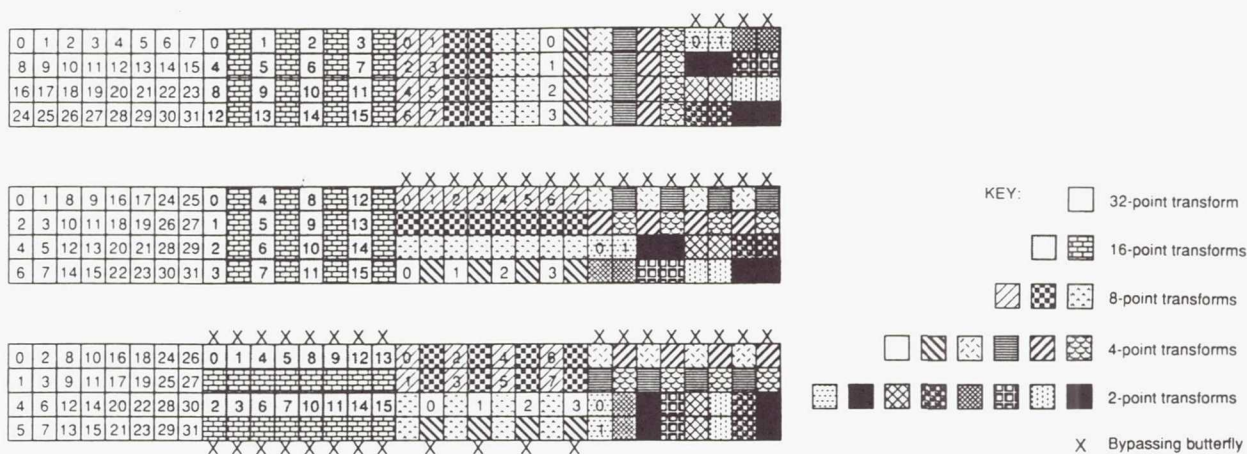
Figure 7. Mixed-Size IFFT

175

Figure 8. Interpolation Filter Control
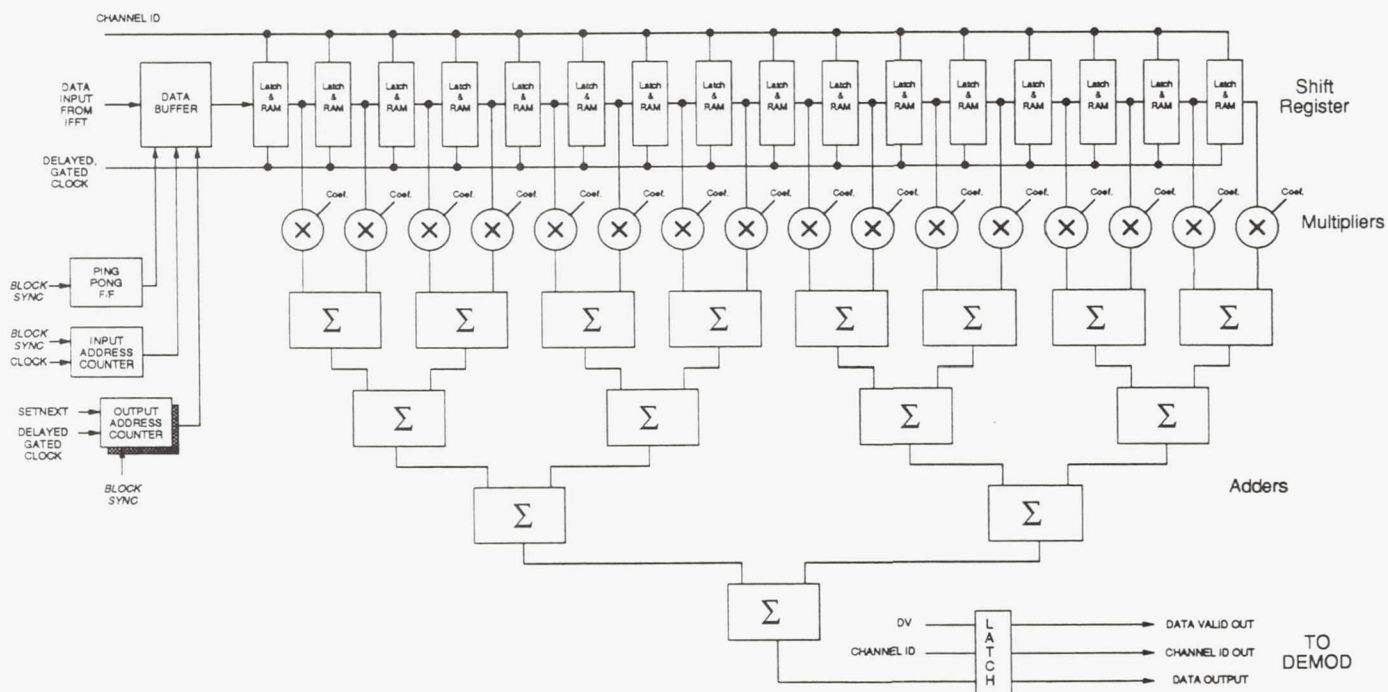
Figure 9. Interpolation Filter Computations
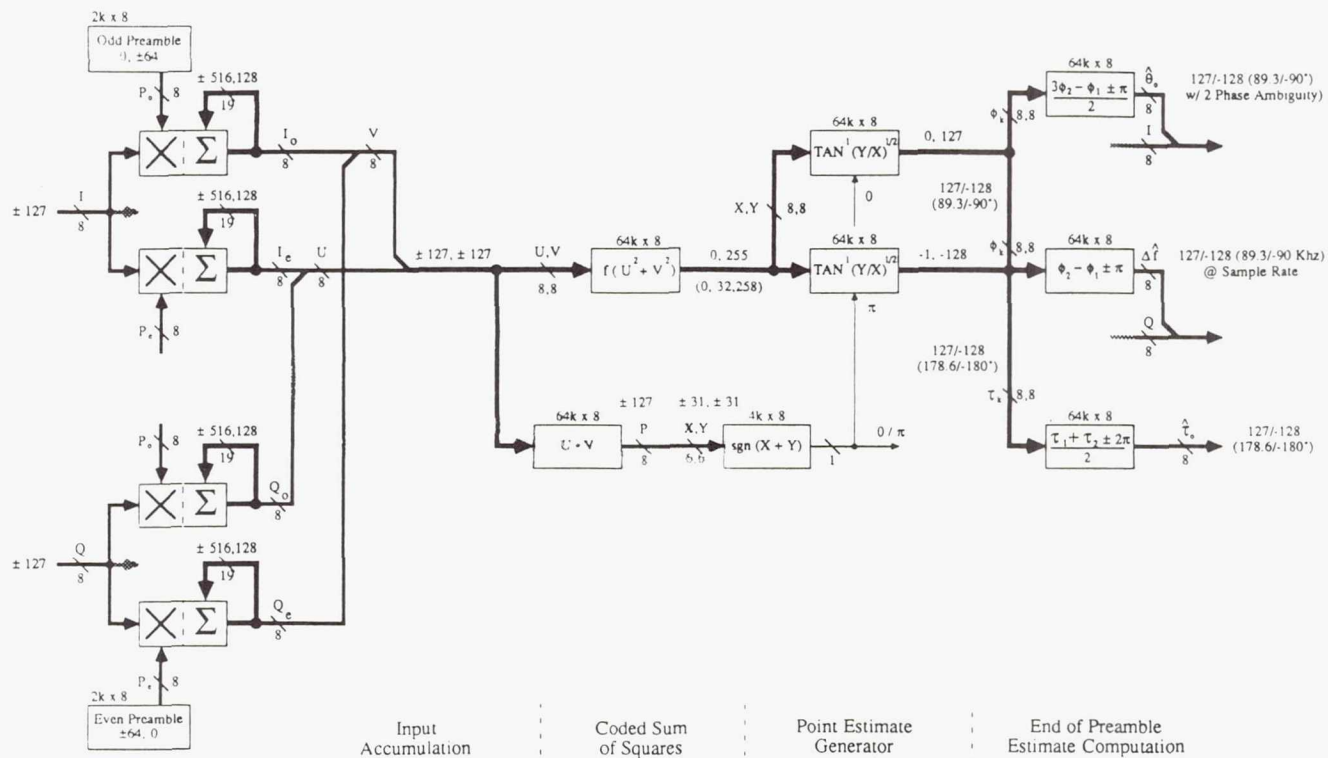
Figure 10. Demodulator Acquisition Control Module
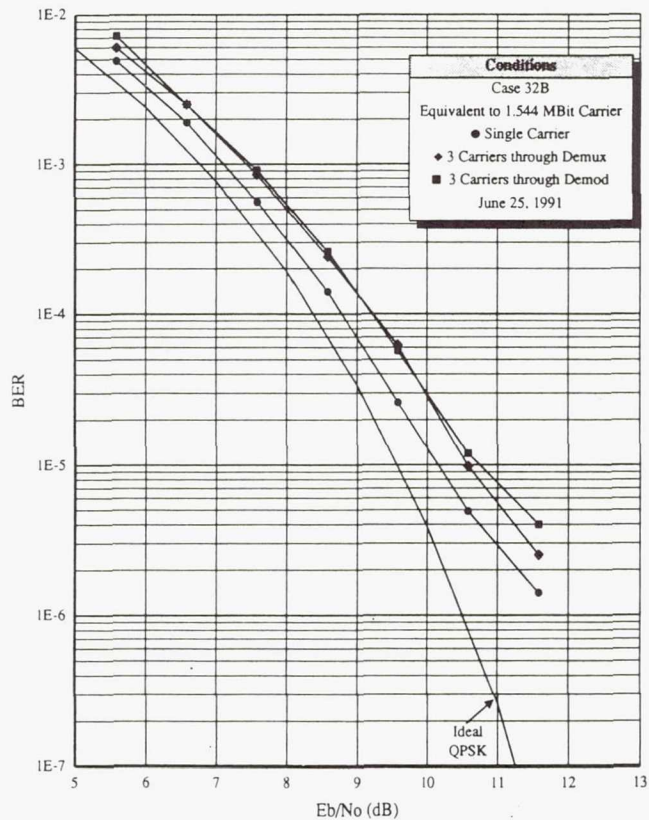
Figure 11. BER Performance

177

Page intentionally left blank

OPTIMIZATION OF AN OPTICALLY IMPLEMENTED ON-BOARD FDMA DEMULTIPLEXER

J. Fargnoli and L. Riddle

Westinghouse Electric Corp.

Baltimore, Maryland

SUMMARY

Performance of a 30 GHz FDMA uplink to a processing satellite is modelled for the case where the on-board demultiplexer is implemented optically. Included in the perforance model are the effects of adjacent channel interference, intersymbol interference, and spurious signals associated with the optical implementation. Demultiplexer parameters are optimized to provide the minimum bit error probability at a given bandwidth efficiency when filtered QPSK modulation is employed.

INTRODUCTION

Satellite communication using frequency division multiple access (FDMA) on the uplinks and time division multiple access (TDMA) on the downlinks has attracted much interest [1,2]. FDMA on the uplink permits the use of ground transmitters that do not require amplifiers having excessively high power. Also, FDMA does not require complicated network timing. TDMA on the downlink takes advantage of recent developments in satellite on-board processing and switching capabilities to provide high data rate downlinks to VSAT-type ground receivers. In addition, the heavily-used C-band and Ku-band frequencies will be supplemented by higher frequency Ka-band transmission (30 GHz uplink / 20 GHz downlink). This permits the use of smaller ground terminal antennas, but at a cost of higher rain attenuation.

On-board processing is needed to efficiently service multiple users while at the same time minimizing earth station complexity. Figure 1 is a simplified overview of a SATCOM system that services FDMA uplink users. The processing satellite first receives the wideband uplink at 30 GHz and downconverts it to a suitable IF. A demultiplexer then separates the composite IF signal into assigned channels. All channels are then demodulated by "bulk" demodulators, with the baseband signals being routed to the downlink processor for retransmission to the receiving earth stations via a high-rate TDMA 20 GHz downlink. This type of processing circumvents many of the difficulties associated with bent-pipe repeaters. First, uplink signal distortion and interference are not retransmitted on the downlink. Second, downlink power can be allocated in accordance with user needs, independent of uplink transmissions. This allows the uplink users to employ different data rates as well as different modulation and coding schemes. In addition, all downlink users will then have a common frequency standard and symbol clock on the satellite, which is useful for network synchronization.
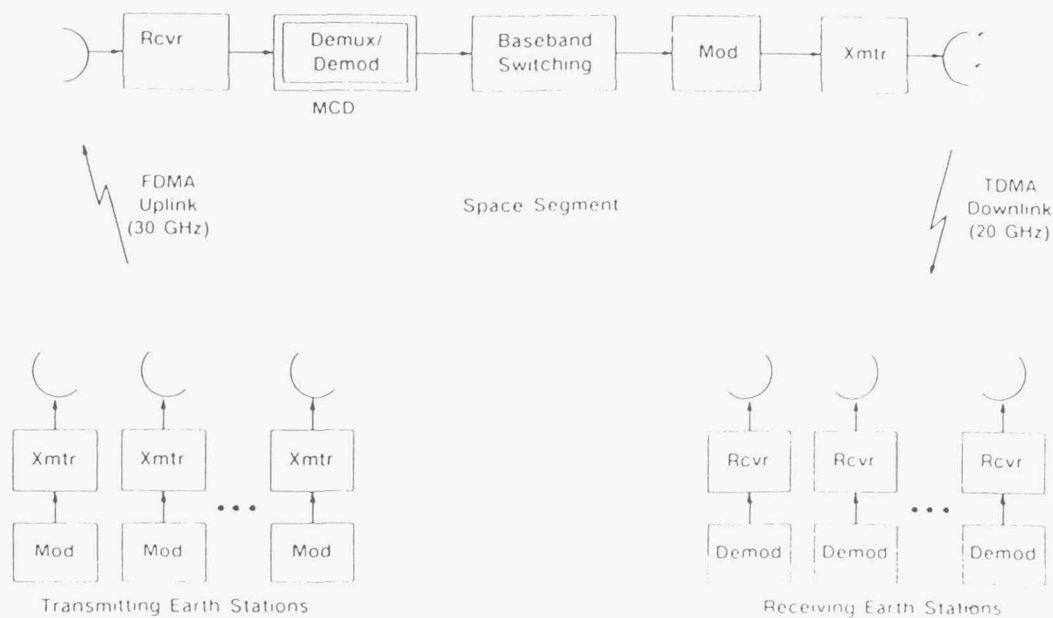
Fig. 1 System Overview

These considerations led to a requirement for on-board multi-channel demodulators (MCD) that can separate and process the individual transmissions with minimal degradation in bit error probability. Implementation of an MCD is critical because future systems will be highly bandwidth-efficient, which implies very close spacing of the carriers in the composite FDMA uplink.

On-board FDMA demultiplexers can be implemented in a variety of ways. One way is to do a wideband A/D conversion on the uplink signal received at the satellite, followed by digital processing that performs the channel filtering and demodulation operations [3]. However, on-board demultiplexing can also be performed using integrated optics [4,5]. An acousto-optical spectrum analyzer performs both down-conversion and channel filtering, with potential savings in hardware size and weight.

This paper shows how an acousto-optical demultiplexer can be modelled in system performance analyses. Bit error performance is determined in the presence of adjacent channel interference, intersymbol interference, and spurious signals generated by the optical processing.

ON-BOARD DEMULTIPLEXER

An acousto-optical spectrum analyzer (Fig. 2) employing heterodyne detection can function as a channelized receiver. The spectrum analyzer converts the composite FDMA uplink into acoustic waves in a Bragg cell. These acoustic waves modulate a laser beam, and diffract the beam at angles proportional to the uplink RF signal frequencies. Reference beams are also provided to achieve heterodyne operation, resulting in larger dynamic range. The diffracted light impinges on an array of photodetectors, which function as square-law detectors, and the individual photocurrents are routed to QPSK demodulators. Thus, the acousto-optical spectrum analyzer serves as both a channelizer and downconverter, so that

180
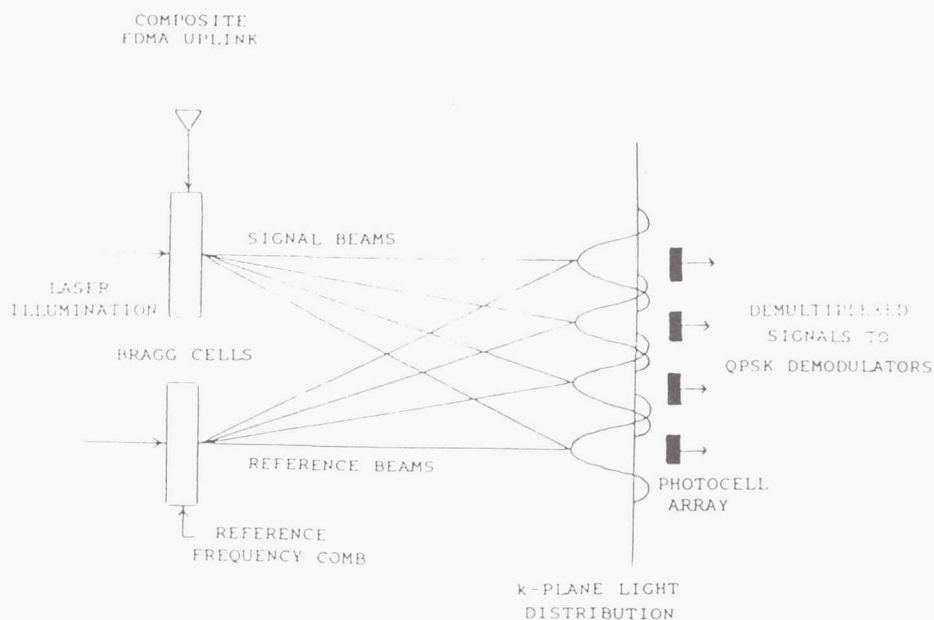
Fig. 2 Acousto-optical Heterodyne Spectrum Analyzer

the composite uplink signal is demultiplexed into separate channels, each at a common IF.

To estimate system performance from a demultiplexer of this type, it is first necessary to determine its transfer function. In a classical linear system, a sinusoidal input to the system results in a sinusoidal output at the same frequency, whose amplitude and phase depend on the frequency. But for the heterodyne system considered here, a sinusoidal input results in not only a sinusoidal output at that frequency, but also sinusoids at $f + \sum nF$, where F is the channel spacing. These spurious sinusoids are generated internal to the demultiplexer by the reference frequency comb. In contrast, the frequencies provided by other transmitters external to the demultiplexer make up the ACI, which is characteristic of all FDMA systems.

DEMULTIPLEXER TRANSFER FUNCTION

As indicated in Fig. 2, two channels constitute the acousto-optical spectrum analyzer: the "signal" channel and the "reference" channel. When the beam has a Gaussian cross section, the light into the signal channel Bragg cell can be expressed as

$$e^{-c_S z^2 + i2\pi f_L t} \tag{1}$$

181

where f_L is the light frequency, x denotes distance from the center of the Bragg cell, and c_S is a constant determined by the laser beamwidth. This light is modulated by an acoustic wave produced by the input sinusoid of frequency f_S, which can be expressed as

$$e^{i2\pi f_S(t-\frac{x}{v})} \tag{2}$$

where v is the acoustic velocity in the Bragg cell. Therefore, the modulated light out of the cell is the product of (1) and (2), and the signal channel light distribution in the k-plane is the spatial Fourier transform:

$$F_S(k) = e^{i2\pi(f_L+f_S)t} \int_{-\frac{d_S}{2}}^{\frac{d_S}{2}} e^{-c_S x^2 - i2\pi(\frac{f_S}{v}+k)x} dx \tag{3}$$

where d_S is the length of the Bragg cell. Similarly, the light distribution resulting from the multi-diffracted beam in the reference channel is

$$F_R(k) = \sum_{n=-\infty}^{\infty} e^{i2\pi(f_L+f_R+nF)t} \int_{-\frac{d_R}{2}}^{\frac{d_R}{2}} e^{-c_R x^2 - i2\pi(\frac{f_R+nF}{v}+k)x} dx \tag{4}$$

One of the reference beams is directed toward the signal channel light distribution (3) in the k-plane. However, because the other reference beams also overlap the signal beam to some extent, spurious output signals occur. If a high bandwidth efficiency is required, then the beams must overlap more, which implies a higher level of spurious signals.

The total light intensity in the k-plane is

$$|F_S + F_R|^2 = |F_S|^2 + |F_R|^2 + 2Re F_S F_R^* \tag{5}$$

By suitably filtering the photodetector output, the only important contribution to the output will be the cross product $G(k) = Re\ F_S F_R^*$. The photocurrent is proportional to the integral of the intensity over the photosensitive area:

$$I = \int_{k_0-\frac{K}{2}}^{k_0-\frac{K}{2}} G(k)dk \tag{6}$$

where k_0 is the location of the photocell in the k-plane and K is the width of the photocell. Now let f_0 be the nominal channel carrier frequency, and let $f = f_S - f_R$ represent the input frequency relative to this nominal frequency. Assume that the k-plane location k_0 corresponds to the frequency f_0, so that $k_0 = -f_0/v$. Then the photocell output, translated in frequency to baseband, reduces to

$$I = \sum_{n=-\infty}^{\infty} H_R(f,nF)\cos 2\pi(f-nF)t \tag{7}$$

182

where

$$H_R(f,nf) \quad \int_0^{u_s} e^{\cdots} \int_0^{u_H} e^{\cdots v'} \left[\cos \frac{2\pi}{\nu}(fx + nf'y) \frac{\sin \pi h(x - y)}{x - y} + \right. \tag{8}$$

$$\left. \cos \frac{2\pi}{\nu}(fx - nf'y) \frac{\sin \pi h'(x + y)}{x + y} \right] dx dy \tag{9}$$

The function $H_R(f,0)$ is the transfer function which will be denoted $H(f)$, and which can be calculated by numerical integration after the system parameters have been selected. The terms for which n=0 give the amplitudes of the spurious sinusoids.

Figure 4 is the computed transfer function for a particular set of MCD parameters. The parameters were selected to give the minimum bit error probability in the presence of ACI and ISI when the earth station transmissions are filtered with fifth-order Butterworth filters having time-bandwidth products of 0.5. A bandwidth efficiency of 1.6 bps/Hz has been assumed. We now describe the method used to compute the ACI and ISI, and then the bit error probability itself.

(a) Spurious Signal Distribution from Reference Comb   (b) ACI Distribution from Interfering Transmitters
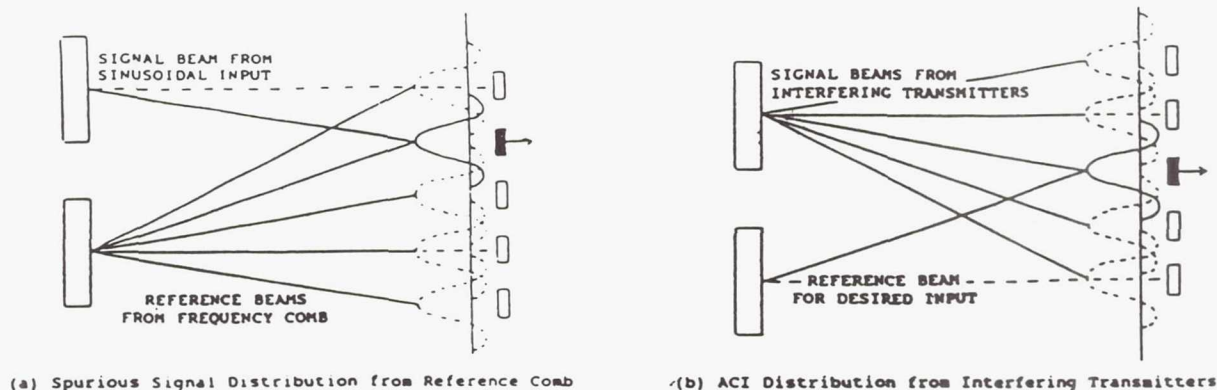
Fig. 3 Interference Effects
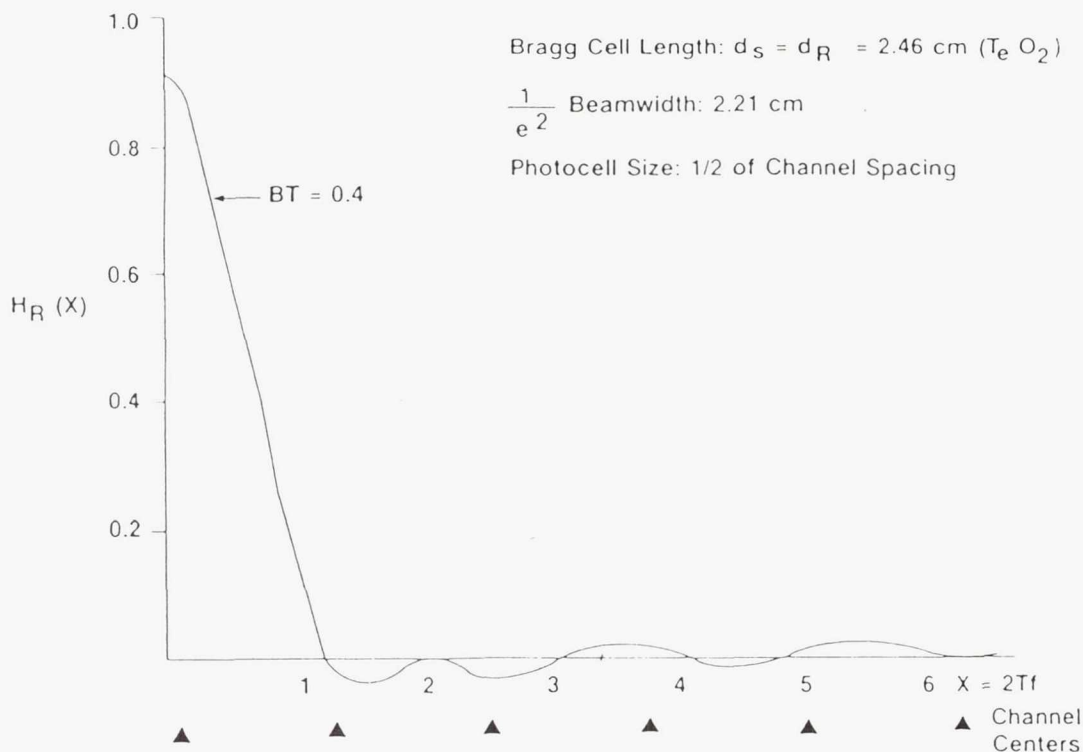
Fig. 4 MCD Transfer Function

ADJACENT CHANNEL INTERFERENCE

Design of any bandwidth-efficient FDMA system involves a fundamental trade-off. If the system bandwidth is narrow, we achieve good ACI performance at the cost of high ISI. Widening the bandwidth reduces the ISI but increases the ACI. The design procedure is generally to select filter types and bandwidths that give the best bit error performance in the presence of both ACI and ISI. We first consider the ACI.

Each ground transmitter is assumed to include bandpass filtering to reduce the amount of ACI entering the satellite receiver. As a first approximation for performance analysis, the demultiplexer can be treated conventionally as a bank of bandpass filters, each followed by a demodulator. The model is later generalized to take into account the spurious signals that are characteristic of the demultiplexer implementation.

It is straightforward to compute the ACI under the assumption that the interference can be treated as noise that adds to the thermal noise at the receiver input. This assumption is valid when there is a large number of interfering users because according to the central limit theorem, this implies that the interference has nearly Gaussian statistics. We also assume for simplicity that all transmissions arrive at the satellite with equal power.

184

The unfiltered spectral density of the n-th QPSK-modulated interfering signal is

$$S(f - nF) = \left[\frac{\sin 2\pi T(f - nF)}{2\pi T(f - nF)} \right]^2 \tag{10}$$

All ground transmitters have identical filters, and the n-th filter transfer function is denoted by $H_T(f-nF)$. Thus the filtered transmission from the n-th interferer has a spectral density given by $S(f-nF)|H_T(f-nF)|^2$. Suppose the transfer function of the on-board demultiplexer is $H_R(f)$, which was evaluated in the previous section. Then the spectral density of the interference into the demodulator is

$$\sum_{n \neq 0} S(f - nF)|H_T(f - nF)H_R(f)|^2 \tag{11}$$

Assume that the symbol detector is a filter matched to the undistorted symbol (i.e. an integrate-and-dump detector). Its transfer function is $H_{MF}(f) = \sin(2\pi Tf)/ 2\pi Tf$. Then the total ACI power out of the matched filter relative to the undistorted signal power is

$$I = \sum_{n \neq 0} \int_{-\infty}^{\infty} S(f - nF)|H_T(f - nF)H_R(f)H_{MF}(f)|^2 df \tag{12}$$

This is added to the thermal noise to estimate the error probability.

INTERSYMBOL INTERFERENCE

Unlike the ACI, the ISI cannot be accurately approximated as additive Gaussian noise. Instead, we determine explicitly the effect of the transmitter and receiver filtering on the amplitude of the signal out of the integrate-and-dump detector.

Expressing any one of the unfiltered QPSK signals before transmission as s(t) = m(t)cos(2 f t +), the complex modulation is

$$m(t) = \frac{1}{\sqrt{2}} \sum_{n=-\infty}^{\infty} \left[a_n \text{ rect} \left(\frac{t - 2nT}{2T} \right) + ib_n \text{ rect} \left(\frac{t - (2n - 1)T}{2T} \right) \right] \tag{13}$$

where a and b are binary data on the I- and Q- channels respectively, which are assumed to be offset 1/2-symbol. The modulation spectrum of the filtered and demultiplexed signal at the input to its demodulator is $\widehat{M}(f) = M(f)H_T(f)H_R(f)$, where M(f) is the spectrum of the undistorted QPSK modulation. The Fourier transform of this is the distorted modulation, which we will call $\widehat{m}(t)$, and the output of the I-channel integrate-and-dump detector in the demodulator is

$$\frac{1}{2T} \int_{-T+\tau}^{T+\tau} \widehat{m}(t)dt = T \sum_{n=0}^{\infty} a_{-n} \int_{-\infty}^{\infty} H_T(f)H_R(f) \left(\frac{\sin 2\pi Tf}{2\pi Tf} \right)^2 e^{i2\pi(2nT+\tau)f} df \tag{14}$$

185

where τ is the sampling time relative to the symbol transition. The Q-channel output is similar. Because the filters introduce group delay, the sampling time is generally nonzero. We can split the composite transfer function $H_T(f)H_R(f)$ into its real and imaginary parts denoted by $\alpha(f)$ and $\beta(f)$ respectively. Because α is an even function of f and β is odd, the signal amplitude out of the I-channel of the detector becomes

$$S = 2 \sum_{n=0}^{\infty} a_{-n} \int_0^{\infty} \left(\frac{\sin \pi x}{\pi x} \right)^2 \left[\alpha(x) \cos 2\pi \left(n + \frac{\tau}{2T} \right) x - \beta(x) \sin 2\pi \left(n + \frac{\tau}{2T} \right) x \right] dx \qquad (15)$$

after changing the integration variable. This shows that the detector output includes contributions from not only the desired (n=0) symbol, but from all symbols, which is what is meant by intersymbol interference. The bit error probability can be estimated from either the I-channel or Q-channel output.

By differentiation, we find that the value of   for which the average value of S is a maximum is the solution of

$$\int_0^{\infty} \left(\frac{\sin \pi x}{\pi x} \right)^2 \left[\alpha(x) \sin \frac{\pi \tau x}{T} + \beta(x) \cos \frac{\pi \tau x}{T} \right] dx = 0 \qquad (16)$$

Once the optimum τ is found we can compute the contribution to S from the n=0 symbol and from all important interfering symbols. It has been found that the n = 1,2, 3, and -1 interfering symbols are the important ones in our application.

The bit error probability is then computed by averaging the error probability, conditioned on a particular sequence of these interfering symbols, over all 16 possible sequences of these symbols. The noise in this computation consists of the ACI, which was calculated in the previous section, plus the thermal noise at the input to the satellite receiver. If N_0 is the thermal noise density at the input, the thermal noise power at the output of the integrate-and-dump detector is

$$N_0 \int_{-\infty}^{\infty} |H_R(f)H_{MF}(f)|^2 df \qquad (17)$$

which is added to the ACI in the error probability computation.

RESULTS

Having derived the demultiplexer transfer function, and having calculated the ACI and ISI, it is straightforward to write down an expression for the total power out of the device:

$$P = \sum_n \sum_m \int_{-\infty}^{\infty} S(f - nF)|H_T(f - nF)H_R(f, mF)H_{MF}(f)|^2 df \qquad (18)$$

The first sum is over the number of interfering earth stations, while the second is over the number of spurious signals produced by the reference beams. The m=n=0 term is the desired output; the m=0, n≠0 terms are the ACI; the m≠0, n=0 terms are the spurious outputs when there is only one uplink signal; and the m≠0, n≠0 terms are additional spurious outputs arising from interaction of the interfering channels with the n≠0 reference beams. Of this latter category, the m=n terms dominate, so they are included in the performance calculation. The m≠n interaction terms are illustrated in Fig. 2.

Figure 6 shows the bit error probability when the MCD transfer function is that shown in Fig. 4. Fifth-order Butterworth filters are used in the earth station transmitters, and integrate-and-dump symbol detection is assumed. The effects of ACI, spurious signals, and ISI are all included in the calculation. Also shown for comparison is a case where the optical MCD is replaced by a bank of bandpass fifth-order Butterworth filters whose bandwidths are optimized to give the smallest bit error probability in the presence of the same interference. Figure 6 shows that the performance of an optical MCD compares favorably with that of an MCD implemented electronically.

Figure 7 shows the sensitivity of system performance to timing errors in the symbol detection circuitry. This is evaluated by varying the sampling time about its optimum value when computing the error probability. It is evident that when the timing error is less than approximately ten percent of the symbol duration, its contribution to bit error degradation is insignificant compared to the ACI, spurious signals, and ISI.
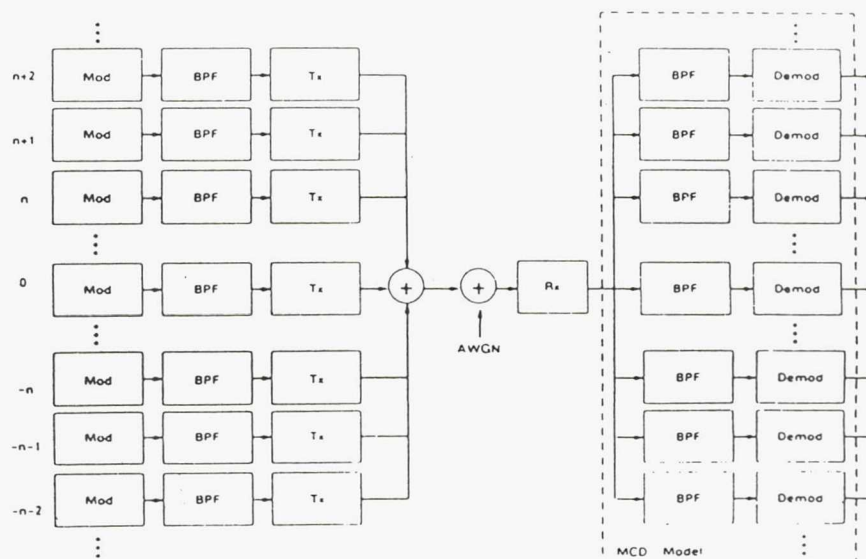
Fig. 5 Analytical Model for Performance Evaluation
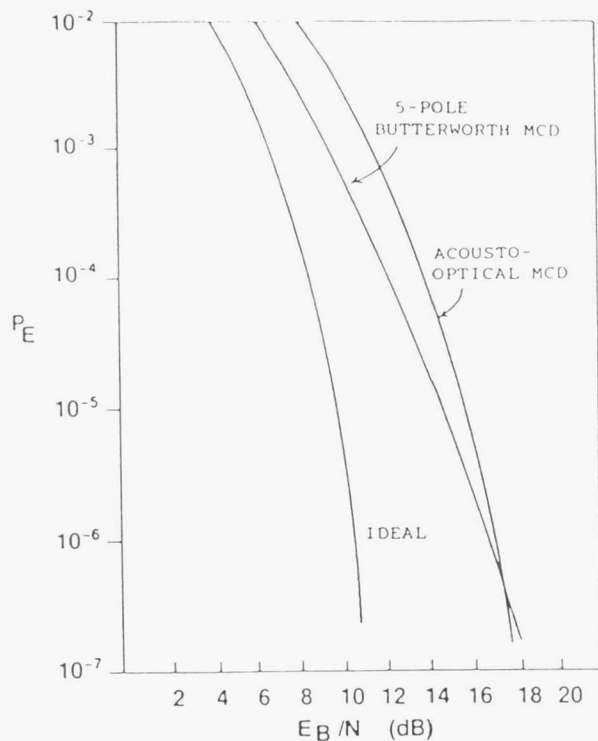
Fig. 6 Bit Error Probability
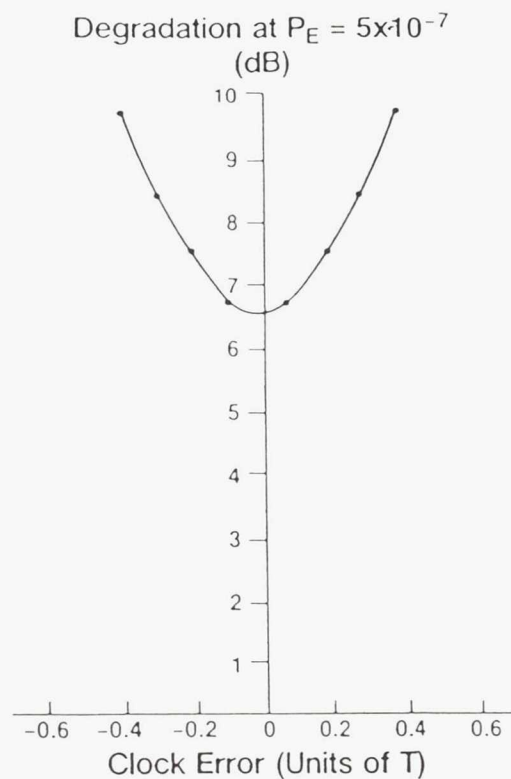
Fig. 7 Sensitivity to Symbol Synchronization Error

DISCUSSION AND CONCLUSION

The goal of this effort was to evaluate the performance of an optically implemented on-board demultiplexer that can service inexpensive, low-power earth stations. For such earth stations, no attempt has been made to improve bandwidth efficiency by using advanced modulation and coding schemes. In addition, we have assumed that no effort has been made to minimize ISI by careful design of the earth station transmitter filters. We have shown that an optically implemented MCD, which promises size and weight advantages over other implementations of on-board processors, can perform as well as other implementations so far as bit error probability is concerned.

However, to achieve comparable performance with the optical MCD, whose transfer function must be carefully controlled, it was found necessary to reduce the ground transmitter filter bandwidths. As seen in Fig. 6, this worsens performance at low signal-to-noise ratios because of the increased ISI that results. The optical MCD transfer function (Fig. 4) has amplitude and phase characteristics different from most classical filter responses. Also, spurious responses from the reference beams exist. Therefore, the ACI/ISI trade-off for an optical implementation differs significantly from that for an electronic one.

188

REFERENCES

[1] Ananasso, F. and Saggese, E., "A survey on the Technology of multi-carrier Demodulators for FDMA/TDMA User-oriented Satellite Systems", IEEE International Conference on Communications, Chicago, 1985, pp. 6.6.1 - 6.1.7.

[2] Ananasso, F. and Del Re, E., "Clock and Carrier Synchronization in FDMA/TDMA User-oriented Satellite Systems", IEEE International Conference on Communications, Seattle, 1987, pp. 41.7.1 - 41.7.5.

[3] Kappes, J. and Sayegh, S., "Programmable Demultiplexer/Demodulator Processor", IEEE Military Communication Conference, Monterey, 1990, pp. 7.3.1 - 7.3.5.

[4] Goutzoulis, A. and Abramovitz, I. "Digital Electronics Meets Its Match", IEEE Spectrum, Aug. 1988, pp. 21-25.

[5] Ananasso, F. and Bennion, I., "Optical Technologies for Signal Processing in Satellite Repeaters", IEEE Communications Magazine, Feb. 1990, pp. 55-64.

Page intentionally left blank

DESIGN, MODELING, AND ANALYSIS OF
MULTI-CHANNEL DEMULTIPLEXER / DEMODULATOR *

David D. Lee and K. T. Woo

TRW Electronic Systems Group
One Space Park
Redondo Beach, CA 90278

I. Introduction

Traditionally, satellites have performed the function of a simple repeater. Newer data distribution satellite architectures, however, require demodulation of many frequency division multiplexed uplink channels by a single demultiplexer/demodulator unit, baseband processing and routing of individual voice/data circuits, and remodulation into time division multiplexed (TDM) downlink carriers. The TRW MCDD (Multichannel Demultiplexer/Multirate Demodulator) operates on a 37.4 MHz composite input signal. Individual channel data rates are either 64 Kbps or 2.048 Mbps. The wideband demultiplexer divides the input signal into 1.44 MHz segments containing either a single 2.048 Mbps channel or thirty two 64 Kbps channels. In the latter case, the narrowband demultiplexer further divides the single 1.44 MHz wideband channel into thirty two 45 KHz narrowband channels.

With this approach the time domain FFT channelizer processing capacity is matched well to the bandwidth and number of channels to be demultiplexed. By using a multirate demodulator fewer demodulators are required while achieving greater flexibility. Each demodulator can process a wideband channel or thirty two narrowband channels. Either all wideband channels, a mixture of wideband and narrowband channels, or all narrowband channels can be demodulated. The multirate demodulator approach also has lower nonrecurring costs since only one design and development effort is needed.

TRW has developed a POC (Proof of Concept) model which fully demonstrates the signal processing functions of MCDD. It is capable of processing either three 2.048 Mbps channels or two 2.048 Mbps channels and thirty two 64 Kbps channels. An overview of important MCDD system engineering issues is presented as well as discussion on some of the BOSS (Block Oriented System Simulation) analyses performed for design verification and selection of operational parameters of the POC model. System engineering analysis of the POC model confirmed that the MCDD concepts are not only achievable but also balances the joint goals of minimizing on-board complexity and cost of ground equipments while retaining the flexibility needed to meet a wide range of system requirements.

II. MCDD Concepts and Architecture

1. MCDD Frequency Plan

The TRW MCDD concept embodies a multistage approach to accommodate different channel bandwidths. Each MCDD operates on a digitized 37.4 MHz bandwidth segment. The initial stage provides channelization into thirty two wideband 1.44 MHz channels corresponding to the wideband data rate of 2.048 Mbps. The signal at this point is either demodulated directly or fed to a second channelizer which subdivides the signal into thirty two 45 KHz channels with 64 Kbps narrowband data rate. In the latter case, the output of the second channelizer is fed to a demodulator which is identical to that used for the wideband channel. This single demodulator recovers all of the thirty two 64 Kbps data streams. Thus, only a single demodulator needs to be provided for each 1.44 MHz subsegment of the input spectrum regardless of whether the subsegment contains a single 2.048 Mbps channel or thirty two 64 Kbps channels.

The WB channel data rate and channel spacing are chosen to maximize the composite FDM signal bandwidth efficiency, and the same ratio of maximum bit rate to channel spacing is maintained for the narrowband
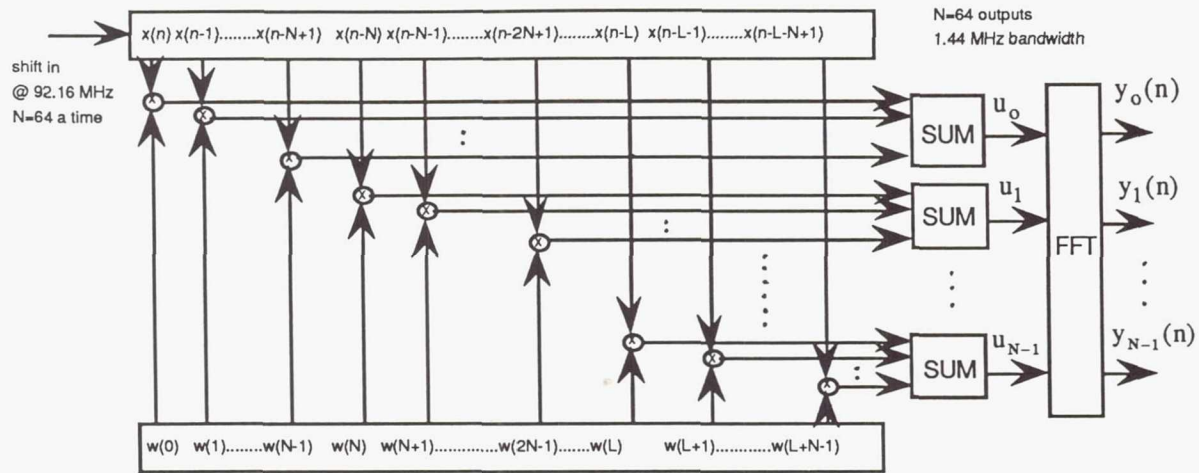

Figure 1. Time Domain FFT Channelizer

channels. Smaller channel spacings place undue frequency and Doppler compensation requirements on the user ground terminals. Larger channel spacings are not only inefficient, but make the typical frequency generation components in the ground terminals unusable. The rate and channel spacing enables a very efficient algorithm to be used within the MCDD for demultiplexing.

2. The Channelizer Design

The TRW MCDD concept uses time domain FFT channelizing in a wideband/narrowband cascade. This algorithm processes the data in the time domain by performing window/presum followed by a single FFT to perform heterodyning of the signal. The FFT channelizer algorithm is important for its efficient use of FFT hardware to realize a parallel bank of filters.

Suppose we require a bank of N complex bandpass filters with equally spaced center frequencies from [DC to f_s), the impulse response of the kth filter $h_k(n)$ can be written in terms of a complex modulated lowpass filter $w(n)$

$$h_k(n) = w(n)e^{-j\left(\frac{2\pi}{N}\right)kn}$$

; k=0,..N-1   (1)

The Fourier transform shows that the frequency response of the kth filter is a frequency shifted version of lowpass filter's frequency response;

$$H_k(e^{j\omega}) = W\left(e^{j\left(\omega+\frac{2\pi}{N}k\right)}\right)$$

; k=0,...,N-1   (2)

Thus, the kth output $y_k(n)$ is given by the convolution

$$y_k(n) = \sum_{m=-\infty}^{\infty} x(n-m)w(m)e^{-j\left(\frac{2\pi}{N}km\right)}$$

;k=0,...,N-1   (3)

By making substitution : m=Nr+q with q=0,...,N-1 and $-\infty \leq r \leq \infty$, we can express equation (3) as a double sum

$$y_k(n) = \sum_{q=0}^{N-1}\left[\sum_{r=\infty}^{\infty} x(n-Nr-q)w(Nr+q)\right]e^{-j\left(\frac{2\pi}{N}kq\right)}$$

;k=0,...,N-1   (4)

since

$$e^{-j\left(\frac{2\pi}{N}kq\right)} = e^{-j\left(\frac{2\pi}{N}k(Nr+q)\right)}.$$

Equation (4) can be rewritten as

$$y_k(n) = \sum_{q=0}^{N-1}[U_n(q)]e^{-j\left(\frac{2\pi}{N}kq\right)}$$

; k=0,...,N-1   (5)

192
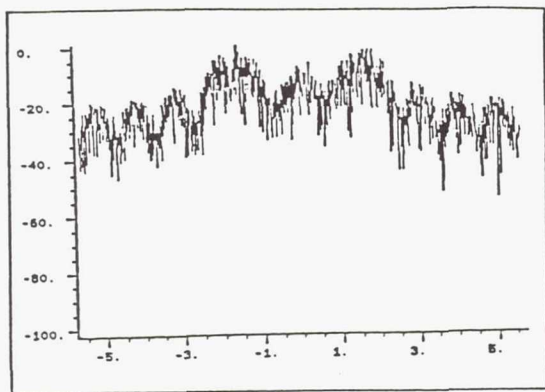
(a) Without Pulse Shaping
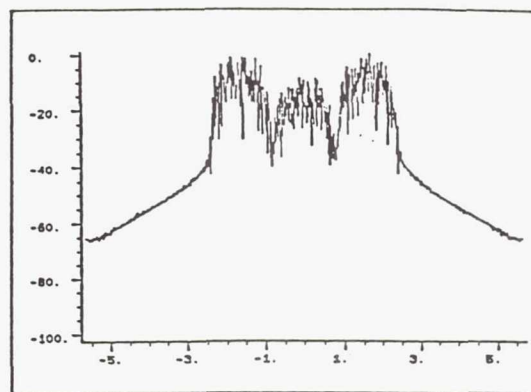
(b) With Baseband Pulse Shaping

Figure 2. Power Spectrum of Three Adjacent SQPSK signals when the center
channel signal power is 8 dB lower than those of two adjacent channels

where

$$u_n(q) = \sum_{r=-}^{-} x(n - Nr - q)w(Nr + q)$$

    ; q=0,...,N-1   (6)

Equation (6) is called the presum operation and becomes a finite sum for finite window length. Figure 1 shows a block diagram of the time domain FFT channelizer derived above.

3. Channel Modulation & Pulse Shaping

SQPSK modulation was considered in the initial phase of the MCDD POC development. Figure 2(a) shows the spectra of three adjacent SQPSK signals when the signal of interest (center channel) has 8 dB less signal power than the other two. BER degradation due to such adjacent channel interference (ACI) is quite significant with SQPSK. To mitigate the effects of intersymbol interference (ISI) and the effects of ACI due to bandlimiting in the channelizer, baseband pulse shaping is applied to the SQPSK I and Q symbol pulses. In particular, the pulse shaping function is a raised cosine response with the roll-off factor of 0.3 equally split between the transmitter and the receiver matched filter.

On the other hand, figure 2(b) shows the spectra of the above SQPSK signals with pulse shaping. The effect of ACI is evidently reduced as compared to that of the SQPSK without pulse shaping. This type of baseband pulse shaping, however, will result in a modulated signal which will no longer be constant-envelope. Similar to bandlimiting filtered SQPSK, the transmitter HPA has to be backed off to its linear range in order not to spread the amplified output spectra. Sidelobe level is a function of the amplifier backoff. For an approximately linear mode of operation of a SQPSK transmitter, a 5 dB output backoff would be required, which significantly reduces the transmitter efficiency. To improve the transmitter power efficiency, a constant envelope bandwidth efficient modulation scheme, i.e., Tamed Frequency Modulation, will be required for MCDD in the future. This modulation type can be demodulated with current MCDD architecture. Our 45 KHz narrowband channel spacing matches that of one of the INTELSAT single-channel-per-carrier services which utilize linearly amplified QPSK.

4. The Demodulator Design

Two types of demodulator processing are performed. Narrowband channels are processed by a time domain narrowband FFT channelizer followed by the demodulator, while wideband channels are directly routed to the demodulator bypassing the narrowband channelizer. The narrowband channelizer is conceptually a scaled-down version of the wideband channelizer. After the channelizer separates the FDM signal, it is processed by the multirate demodulator programmed for narrowband operation. A feature of the demodulation is that the same hardware performs both narrowband and wideband functions with a control input determining the operating mode.

The multirate demodulator consists of five major functional blocks: a *resampling filter* to generate exactly two complex signal samples per symbol from the channelizer output of 1.406 samples per symbol and also matched filter the transmitted signal, a *derotate complex multiply* to remove residual carrier phase error from these samples, a *symbol decision* to determine the received two-bit symbol; a *carrier tracking loop* to estimate the carrier
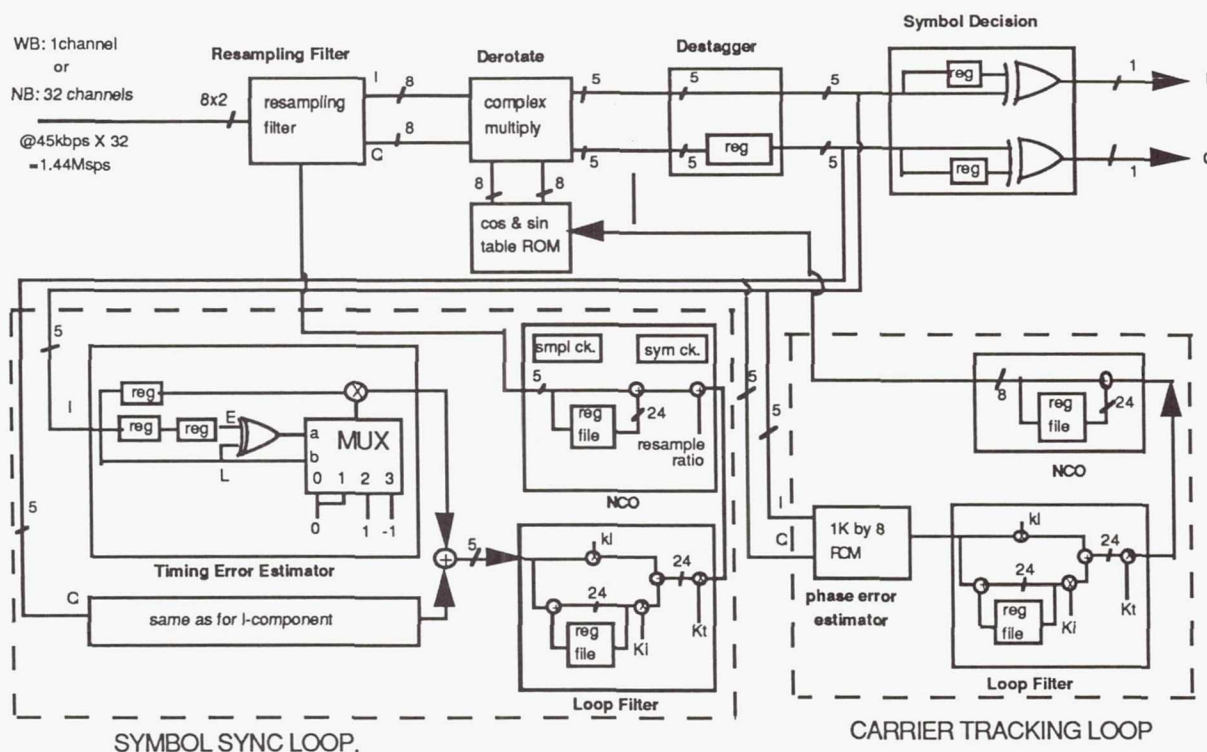
Figure 3. Functional Diagram of the Demodulator

offset to be removed, and a *symbol sync loop* to determine any timing offset to be removed by the resampling filter. Detailed block diagram of the demodulator is shown in Figure 3.

Resampling Filter

The main function of the resampling filter is to resample the data sequence provided by the channelizer at the appropriate sample epochs for the demodulator. The data rate from the channelizer is determined by the ratio of the channel spacing to the symbol rate. Since the narrowband channel spacing is 45 KHz (1.44 MHz for wideband) and the symbol rate is 32 Ksym/sec (2.048 Mbps for wideband), the channelizer output rate is 1.406 samples per symbol. However, the demodulator operation requires 2 samples per symbol for proper timing recovery. Thus the resampler must provide 1.422 samples to the demodulator for every sample provided by the channelizer on the average.

The resampling filter in the MCDD demodulator utilizes the principle of sampling theorem, namely, that a band-limited function is completely determined by its sample values taken at or faster than the Nyquist sampling rate. The reconstructed signal is a summation of appropriately delayed (sin*x*)/*x* functions whose amplitude at the associated sampling instant is exactly the sample value and at every other sampling instant is exactly zero. Furthermore, at all intermediate points in time, the entire collection of terms combines to yield exactly the original continuous waveform everywhere.

In order to do this, the resampling filter works in conjunction with the NCO of the symbol sync loop to interpolate the samples at the correct timing epochs. In order to reduce complexity, the interpolator range is restricted to span only one channelizer sample. Thirty two fractional timing phases are provided by a bank of filter taps. The NCO phase selects the appropriate set for each output sample. Sometimes the resampling filter will compute two output samples for demodulation from a single input sample from the channelizer, and sometimes it will compute only one. (In order for the resampling filter to compute either zero or three samples from a single input sample would require the symbol rate to be off by a degenerate amount.)

194

Derotator

Because of frequency offsets, Doppler shifts, and phase noise, the complex signal at the output of the channelizer and the resampling filter is not be a true SQPSK baseband signal. That is, the signal at the derotator input is given by

$$\tilde{H}_r(k) = \tilde{H}_b(k)e^{j\theta(k)}$$
(7)

where $\tilde{H}_b(k)$ is the true QPSK baseband signal, and $\theta(k)$ is extraneous phase modulation. Because of this extraneous phase modulation, the signal is best described as being a pseudo-baseband signal. The function of the carrier recovery loop is to track the extraneous phase modulation and provide a phase estimate to the derotator for each sample out of the resampling filter. The function of the derotator is to use the phase estimate to translate the pseudo-baseband signal back to the baseband. In order to do this, the derotator multiplies the each input sample by the unit-valued phasor which has a phase in the opposite direction of the phase estimate. The derotator output at time k is thus given by

$$\tilde{H}_d(k) = \tilde{H}_r(k)e^{-j\hat{\theta}(k)} = \tilde{H}_b(k)e^{j(\theta(k)-\hat{\theta}(k))}$$
(8)

where $\hat{\theta}$ is the phase estimate provided by the carrier tracking loop at time k, and $\theta(k)-\hat{\theta}(k)$ is the residual phase noise. Appropriate design of the carrier tracking loop produces a residual phase noise process with an acceptable RMS value. The implementation of the derotator in integer arithmetic is straight forward.

Carrier Tracking Loop

The carrier phase error estimate taken after the destagger module is low-pass filtered and used to drive a NCO which tracks the carrier phase offset. For computing the loop parameters as shown in figure 3, the following formulae are used:

$$K_i = \frac{2\xi\omega_n T}{AK_v K_i T}$$
(9)

$$K_i = \frac{(\omega_n T)^2}{AK_v K_i T}$$
(10)

where A is the carrier phase estimator module gain, K_v is the NCO gain, T is the symbol period, ξ is the damping factor ($\xi=0.707$), w_n is the loop resonance frequency given by

$$\omega_n = 1.89B_L,$$
(11)

and B_L is the desired loop bandwidth.

With loop bandwidth selection of 10^{-2} times the symbol rate ($B_L T=10^{-2}$), the carrier loop will be able to acquire frequency offsets of 102.4 KHz for the wideband channel, and 3.2 KHz for the narrowband channel, both assuming K-band (14 GHz) uplinks. This implies oscillator instabilities of approximately 7.3×10^{-6} and 2.3×10^{-7} for the wideband and narrowband transmitters, respectively. Thus for wideband channels, it is not necessary to provide any special acquisition mechanism, but an acquisition aid may be required for narrowband channels.

Narrower loop bandwidth decreases RMS jitter due to thermal noise, however, this results in slow pull-in time and poorer phase noise performance due to oscillator instability. On the other hand, setting the loop bandwidth too large would allow too much RMS jitter due to thermal noise which results in substantial BER performance degradation. The loop bandwidth in the MCDD POC model is selected to be one hundredth of the data rate which results in phase jitter due to thermal noise less than six degrees and pull in time of less than 0.05 second.

Symbol Synchronization Loop

The derotated complex samples are processed, I and Q separately, to determine timing error. If the two mid-symbol samples are of opposite signs, they will cancel and the transition sample will be near zero. The value of transition sample is therefore zero if the resampling filter is precisely locked to the symbols, and any timing error will be proportional to the non-zero sample value at the transition. The difference in sign of the two mid-symbol
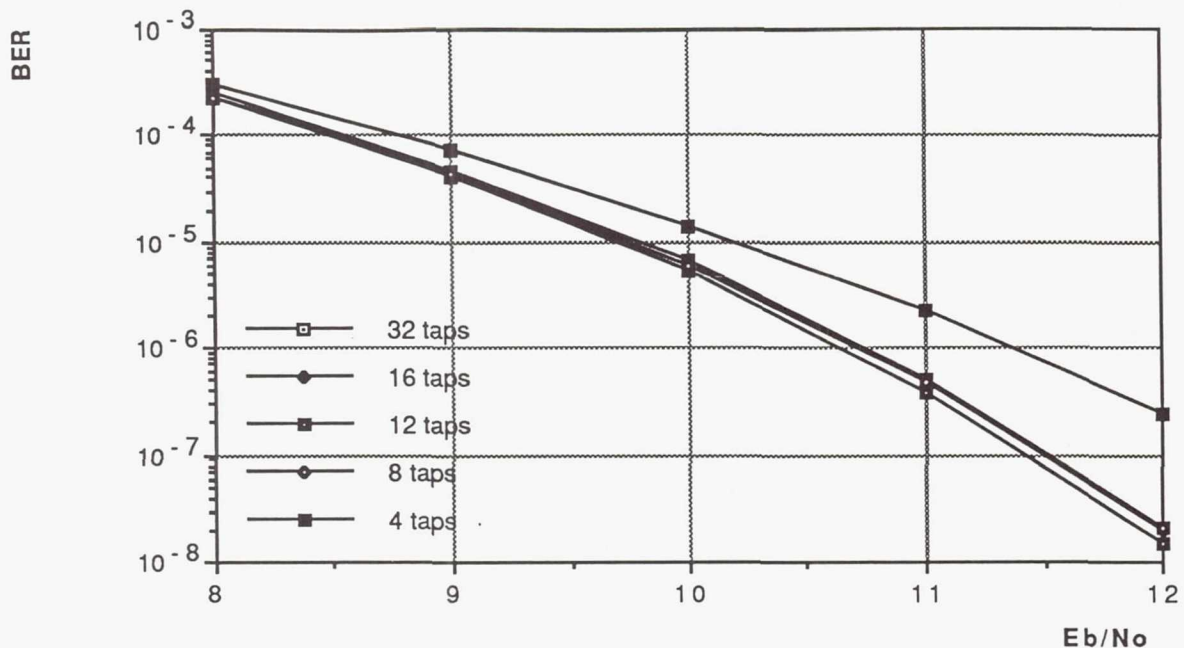
195

Figure 4. BER vs. No. of Resampling Filter Taps

samples indicates the direction of timing error. If the signs are equal, the above sum is ignored since no transition occurred. The measured timing errors for the I and Q samples are summed to give the resulting timing error estimate.

The timing error estimate is low-pass filtered and used to fine tune the NCO which tracks the timing phase and selects the resampling filter taps. The step size is a function of the fixed ratio of the resampling filter input and output sample rates. The fine tuning tracks variable timing drift. The timing offset is output to the resampling filter to select the next filter to be used. The number of resampling filter taps is chosen to provide good interpolation precision while minimizing the size of the hardware. The loop parameters can be obtained using the equations (9) to (11) as well. With the loop bandwidth selection of 5×10^{-3} times the symbol rate, the bit synchronization loop will be able to acquire clock rate offsets of 51.2 Kbps for the wideband channel, and 1.6 Kbps for the narrowband channel. This implies symbol clock instabilities of 2.5×10^{-2} for both the wideband and narrowband transmitters, thus it is not necessary to provide any special acquisition mechanism in either the wideband or the narrowband case.

III. MCDD Performance Analysis Using BOSS

The Block Oriented Systems Simulator (BOSS) run under VMS operating system on a DEC VAX station provides a complete iterative environment for simulation-based analysis and design of any system signal processing operation. Whereas BOSS can perform a time domain (waveform level) simulation of any system, the current model library contains functional blocks most suitable for communication systems simulation. Performance of the total MCDD as well as each of the individual blocks have been carefully analyzed using BOSS. Furthermore, the BOSS semi-analytic BER estimator module was used to perform the parameter trade study. Shown below are two of the most significant parameter trades performed using BOSS.

The demodulator losses due to resampling filter, carrier tracking loop, bit synchronization loop, and quantization, all in terms of the Eb/No degradation from theory for achieving 5×10^{-7} BER, have been quantified. It was found that most of the performance losses came from the resampling filter. Among many resampling filter parameters, the number of resampling filter taps was found to be the most critical to BER performance. Figure 4 shows the BER plots (with ideal channelizer) *vs.* the number of resampling filter taps. Even though eight seems to be a point of diminishing return, our choice for the number of resampling filter taps is sixteen in order to allow room for implementation losses, and future constant envelope modulation choices, i.e., Tamed Frequency Modulation, which will require the longer filter span.
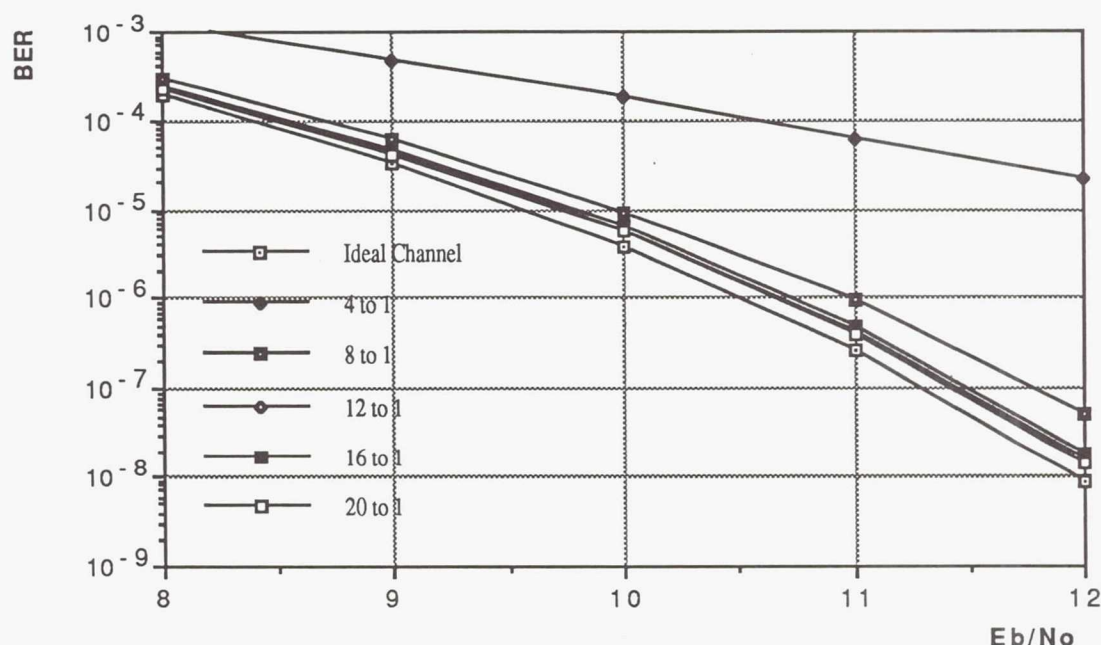
196

Figure 5. Total BER vs. Channelizer Presum Ratio

The channelizer performance principally depends on the window type and presum ratio. Kaiser window has been chosen for its flat pass-band response at the expense of moderate stopband ripple since the current modulation choice with baseband pulse shaping have already mitigated the effects of ACI. The BER performance improves with increasing presum ratio due to wider 3 dB bandwidth and steeper cutoff. It reaches to the point of diminishing return at around twelve. (see figure 5.) Our choice of the presum ratio is sixteen, however, to allow room for additional implementation losses as well as the possibility of implementing a constant envelope bandwidth efficient type modulation later. In the same manner, we have also chosen the window tap quantization to be sixteen bits after observing significant degradation of BER performance with less than twelve bits.

Raised cosine pulse shaping effectively mitigates the effects of ACI and ISI. The worst case ACI example shown in figure 2(b) only increases the BER by less than one percent over the case without ACI. Total MCDD BER estimates have been obtained using the chosen parameters and linearly amplified SQPSK modulation. Note the 0.25 dB Eb/No degradation at the BER of 5×10^{-7} with the presum ratio of 16 to 1 (see figure 5). This compares with about 0.1 dB Eb/No degradation due to the demodulator alone.

IV. Conclusions

A overview of important MCDD system engineering issues have been presented as well as discussion on some of the BOSS analyses performed for design verification and selection of operational parameters of the POC model. The bandwidth-efficient FDM composite signal structure, linearly amplified SQPSK modulation, and time domain FFT channelization method were chosen based on the TRW's space communication system engineering experience and trade studies. Virtually no constraints are imposed by this FDM composite signal structure, which is feasibly implementable on-board due to the efficient wideband and narrowband channelization cascade. The coherent SQPSK demodulator, implementable on-board in advanced VLSI, offers robust operation with imperfect user synchronization.

Acknowledgments

Page intentionally left blank

1

APPLICATION OF CONVOLVE-MULTIPLY-CONVOLVE SAW PROCESSOR FOR SATELLITE COMMUNICATIONS*

Y.S. Lie and M. Ching
Amerasia Technology
Westlake Village, CA

ABSTRACT

There is a need for a satellite communications receiver that can perform simultaneous multi-channel processing of single channel per carrier (SCPC) signals originating from various small (mobile or fixed) earth stations. The number of ground users can be as many as 1000. Conventional techniques of simultaneously processing these signals is by employing as many RF-bandpass filters as the number of channels. Consequently, such approach would result in a bulky receiver, which becomes impractical for satellite applications. A unique approach utilizing realtime surface acoustic wave (SAW) chirp transform processor is presented. The application of a Convolve-Multiply-Convolve (CMC) chirp transform processor is described. The CMC processor transforms each input channel into a unique timeslot, while preserving its modulation content (in this case QPSK). Subsequently, each channel is individually demodulated without the need of input channel filters. Circuit complexity is significantly reduced, because the output frequency of the CMC processor is common for all input channel frequencies. The results of theoretical analysis and experimental results are in good agreement.

I. INTRODUCTION

On-board processing of digital communications signals is necessary to maintain good performance quality of satellite repeaters. In the case of SCPC transmissions the necessity for a large number of analog filters can be avoided by utilizing a Fourier transform processor. Realtime Fourier transforms can be obtained via SAW chirp transformers. In this system, each channel is transformed into a timeslot at the output of the processor. These timeslots are organized in a frame defined by the chirp period of the processor. Figure 1 describes the timing architecture of a QPSK receiver, where detection of two channels are shown. It is clear that the entire network must be synchronized in order for the system to function. In Figure 2 a functional diagram of the receiver is shown. In this case a CMC processor is used. The advantage of utilizing the CMC processor is its inherent ability to transform each input carrier frequency into the time domain on a common frequency. Consequently, only a single circuit is needed for demodulation. The framelength of the processor is made equal to the symbol period of the QPSK signal, hence each channel processor has that length of time to make symbol decisions. However, if all users are transmitting simultaneously, then the signals at the output of the demodulator are generated at a much faster rate. Therefore, it is necessary to store these signals temporarily during processing.

II. THE CMC CHIRP TRANSFORM PROCESSOR

The CMC chirp transform processor model is shown in Figure 3. The input and output filters have an impulse response with a positive slope (i.e. an up-chirp) as shown,

---

where the rate of the linear frequency change is equal to 2k. The chirp generator produces a linear frequency chirp that has a negative slope equal to -2k. It has been shown elsewhere (Ref. 1,2) that the time-bandwidth product of the input and output filters has to be identical (BT), while the chirp generator time-bandwidth product must be at least four times larger (4BT). Note that the input filter has an inherent characteristic of delaying high frequency signals more than low frequency signals. It will be shown later that the output frequency of the CMC chirp transform is constant (ω_2).

To generalize the analysis, the input signal is represented in complex form, as follows:

$$S_i(t) = e^{j\omega_s t} \quad\text{-------------- (1)}$$

The output of the input filter is delayed by d seconds, where d is dependent on the input frequency ω_s, hence

$$S_i(t\text{-}d) = e^{j\omega_s(t\text{-}d)} \quad\text{----------- (2)}$$
$$d = (\omega_s\text{-}\omega_o)/2k \quad\text{---------------- (3)}$$

In the Multiplier, the product of the delayed input signal and the chirp signal is generated. This product is convolved with the impulse response of the output filter to produce the output signal $S_o(t)$, therefore,

$$S_o(t) = \int_{\text{-}T/2}^{T/2\text{-}\tau} x(t).y(t).h_2(\tau\text{-}t)d(t) \quad\text{------------- (4)}$$

For $0 \leq \tau \leq T/2$

Where, x(t) is given by (2), and y(t) and $h_2(t)$ are as follows,

$$y(t) = \text{Exp}[j(\omega_1 t\text{-}kt^2] \quad\text{----------- (5)}$$
$$h_2(t) = \text{Exp}[j(\omega_2 t + kt^2)] \quad\text{--------- (6)}$$

Since the input signal $S_i(t)$ is complex, the output, $S_o(t)$, of the processor is also complex. In this analysis, only the real part of $S_o(t)$ is considered. It can be shown that,

$$R_e[S_o(t)] = \text{Cos}[\text{-}\omega_s d + \omega_2\tau + k\tau^2 + k\epsilon\tau]x(T\text{-}\tau)(\text{Sinc } q) \quad\text{----- (7)}$$

Where,

$$\text{Sinc } q = (\text{Sin } q)/q$$
$$q = [(\omega_s + \omega_1\text{-}\omega_2\text{-}2k\tau)(T\text{-}\tau)/2]$$

ϵ is defined as,

$$\epsilon = \tau - (\omega_s + \omega_1\text{-}\omega_2)/2k \quad\text{---------- (8)}$$

ω_s, ω_o, ω_1, and ω_2, are as defined in Figure 3.

Based on the above model, a computer analysis is performed to demonstrate the transform output of a CW input signal. The CMC chirp transform parameters used for this purpose are as follows:

Input frequency: $f_s = 100$ MHz, $\omega_s = 2\pi f_s$
Center frequency of input filter: $f_o = 100$ MHz
Center frequency of output filter: $f_2 = 300$ MHz
Center frequency of the chirp: $f_1 = 200$ MHz
Time-Bandwidth product of the input and output filters is: BT = 512
Time-Bandwidth product of the chirp generator is: 4BT = 2048
The chirp slope is: $2k = 4.2\times10^6\pi$ mrad/second

As can be seen from equations (3) and (8), when the input frequency f_s is equal to f_o, $d = 0$, and $\epsilon = \tau$. The computer analysis result is shown in Figure 4. The output frequency is

300 MHz, and the envelope of the waveform is a sinc function given by (7). The width of the mainlobe (null-to-null) is approximately 60 nanoseconds, and the peak of the largest sidelobe is 13.5 dB below the main peak as expected from the sinc characteristic. In a multichannel communications system the sidelobes must be suppressed to prevent adjacent channel interference. For that purpose a weighting function can be designed into the SAW chirp filter. In what follows, sidelobe reduction is demonstrated by utilizing Hamming weighting.

III. THE EFFECT OF HAMMING WEIGHTING

To reduce the sidelobes of the chirp transform output, Hamming weighting can be incorporated in the SAW dispersive filter or in the chirp generator. In practice the basic chirp waveform is generated by a software controlled digital circuit. Therefore, it is easier to incorporate the weighting function in the chirp generator, hence the analysis is done with a weighted chirp waveform from the chirp generator. The generalized Hamming (Ref. 3) weighting function is given by:

$$W(t) = a + (1-a)Cos(\pi t/T) \quad -T \le t \le T \text{ ------ (9)}$$
$$0 \le a \le 1$$
$$W(t) = 0 \quad \text{otherwise}$$

In complex form, the chirp function becomes:

$$y(t) = a.Exp[j(\omega t - kt^2)] + [(1-a)/2]\{Exp\ j[(\omega + \pi/T)t - kt^2] +$$
$$Exp\ j[(\omega t - \pi/T)t - kt^2]\} \text{ ----------------- (10)}$$

The chirp transform output, $S_{oh}(t)$, with Hamming weigthing is obtained by replacing the chirp function $y(t)$ in (4) by (10). By solving the resulting convolution integral equation and taking its real value, the transform output, for $0 \le \tau \le T/2$, is as follows:

$$R_e[S_{oh}(t)] = a(T-\tau)Cos[-\omega_s d + (\omega_2 + k\epsilon)\tau + k\tau^2][Sin\ k\epsilon\ (T-\tau)]/k\epsilon\ (T-\tau) +$$
$$[(1-a)(T-\tau)/2]Cos[-\omega_s d + (\omega_2 - \pi/2T + k\epsilon)\tau + k\tau^2]\ X$$
$$\{Sin[(\pi/T - 2k\epsilon)(T-\tau)/2]\}/[(\pi/T - 2k\epsilon)(T-\tau)/2] +$$
$$[(1-a)(T-\tau)/2]Cos[-\omega_s d + (\omega_2 + \pi/T + k\epsilon)\tau + k\tau^2]\ X$$
$$\{Sin[(\pi/T + 2k\epsilon)(T-\tau)/2]\}/[(\pi/T + 2k\epsilon)(T-\tau)/2] \text{ -------- (11)}$$

Utilizing the above equation performance of the CMC chirp transform is analyzed, and the results shown in Figure 5, for a = .54. As can be seen, the sidelobes are reduced relative to the mainlobe when compared to the result in Figure 4. Note, however, that the mainlobe is also slightly reduced. There are other weighting functions that can produce better sidelobe suppression, e.g. Taylor weighting (Ref. 4).

IV. DIGITAL CHIRP WAVEFORM GENERATOR

In the final system design, recovery of the data will require a chirp waveform of +/-32.8 MHz with a period of 31.250 microseconds. A chirp waveform generator (CWG) was designed to provide the in-phase (I) and quadrature (Q) waveforms for the system. Because the output waveforms are calculated by the CWG's firmware, the chirp characteristics can be easily modified for laboratory tests of the system.

A functional block diagram of the CWG is shown in Figure 6. A microprocessor board communicates with a terminal, host processor, or system controller via RS-232C or IEEE-488 interfaces. Chirp waveform data can either be calculated by the internal microprocessor or downloaded from the host for added flexibility. When the CWG is used with a host or IEEE-488 system controller, the resident EPROM monitor recognizes low

level commands to set the address generator and load data in each of the sixteen fast RAM (25 ns access time) chips. Other commands set the step attenuators, on-board timers, and implement debugging functions.

The address generator receives a clocking signal from the multiplexer circuitry at one-eighth of the high speed clock or 16.384 MHz. Thus adequate time is available for address setup and data access without using extremely fast RAMs. In testing, the CWG has been clocked at 250 MHz with 25 ns RAMs. During synchronous mode operation, the address generator is phase locked to a system timing signal to facilitate data channel separation in the time domain.

Each multiplexer is implemented as a sixty-four bit shift register with successive data bytes interleaved eight at a time via a sixty-four bit input latch. The data is taken out in byte parallel form as the shift register is clocked and applied to the DACs. Each DAC can drive a 37-ohm load or a doubly-terminated 75-ohm line.

Low pass filters with corner frequencies of 50 MHz are included to suppress feedthru of the sampling clock.

V. EXPERIMENTAL RESULTS

To verify the theoretical results described above, an experimental circuit was built using existing dispersive SAW filters. The circuit used for this experiment is shown in Figure 7. The dispersive SAW filters have the following characteristics:

Center frequency: 70 MHz
Bandwidth (40 dB): 20 MHz
Weighting function: Taylor
Chirp slope: .253 MHz/μsecond
Time delay: 80 microseconds

Because both SAW filters operate at the same frequency, frequency conversions have to be incorporated to construct the CMC processor so that its input and output center frequencies are the same. This is accomplished by the two mixers, ZFM-1W, shown in Figure 7. The chirp waveform generator and SSB modulator were developed for the NASA/LEWIS SBIR Contract No. NAS3-25862. The output chirp waveform of the SSB modulator is measured utilizing Hewlett Packard's 400 Msample/second digital oscilloscope, and the result is shown in Figure 8. In this diagram only the center part of the chirp is shown, because of limited resolution of the plotter to cover the entire period of the chirp. The chirp waveform was adjusted to be the same as the SAW filter chirp slope, i.e. .253 MHz/μsecond, and the chirp period is approximately 90 μseconds. Since the Time-bandwidth product of the chirp generator has to be four times the SAW filter's, only the center region of the dispersive filter's passband is used to obtain a proper transform output.

Initially, the input frequency is tuned to 70 MHz and the CMC processor output waveform is plotted as shown in Figure 9. As expected the output frequency is also 70 MHz. Subsequently, the input frequency is changed to 70.0015 MHz, and the output waveform is recorded and shown in Figure 10. Note that the output waveform is shifted in phase by 300°. The significance of this result is that in order to obtain coherent demodulation of phase modulated signals, the carrier phase tracking circuit has to take into account the phase shift introduced by the CMC processor.

It was theoretically established (see equation 11) that the output frequency of the CMC processor is constant, in this case 70 MHz. To test the validity of this theory, the input frequency is arbitrarily changed to 75 MHz, the output waveform of the CMC processor is plotted and shown in Figure 11. By comparing the diagrams in Figures 10 and 11, it can be seen that the two outputs have the same frequency. It is therefore confirmed that a common demodulator can be utilized for all input channels.

Figure 12 shows the center portions of the I and Q baseband chirp waveforms with the parameters given earlier. The filtered outputs are applied to the single sideband modulator.

VI. CONCLUSIONS

The CMC processor reduces the complexity of a satellite borne receiver, by transforming all channels into a common frequency, but in different time slots. Thus only a single demodulator is required for all channels. The chirp transform process has been analyzed and experimentally verified. A software controlled digital chirp waveform generator was developed to replace the conventional SAW chirp generator. This technique provides flexibility to compensate for imperfect chirp slope and also to introduce a weighting function for sidelobe suppression.

VII. REFERENCES

1. Williamson R.C., et.al., "A satellite-borne SAW chirp transform system for uplink demodulation of FSK communication signals", Ultrasonic Symposium, 1979.
2. Otto O.W., "Chirp Transform Signal Processor", Ultrasonic Symposium, 1976.
3. Rabiner L.R., et.al., "Theory and Applications of Digital Signal Processing", p.-91, Prentice-Hall, 1985.
4. Skolnik M.I., "Radar Handbook", Naval Research Laboratory, 1978.
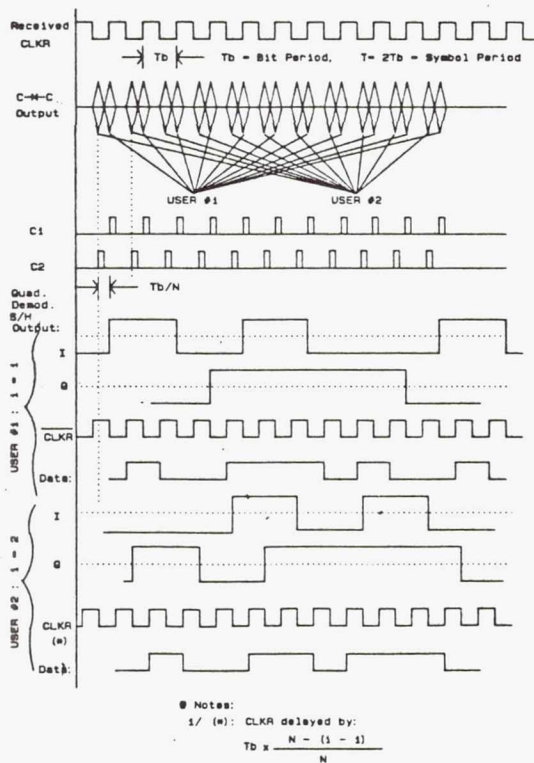
FIGURE 1, SIGNAL AND TIMING ARCHITECTURE OF QPSK RECEIVER
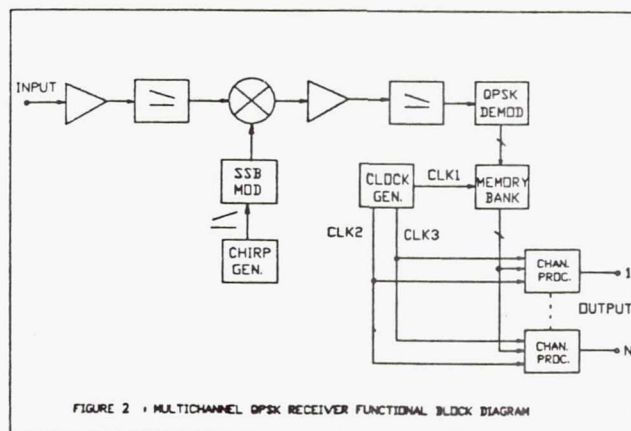
FIGURE 2 : MULTICHANNEL QPSK RECEIVER FUNCTIONAL BLOCK DIAGRAM

TRANSFORM OUTPUT WITH HAMMING

COEFFICIENT SET TO .54

OUTPUT
(Times 10E-5)

2
1.5
1
0.5
0
-0.5
-1
-1.5
-2

-1   -0.8   -0.6   -0.4   -0.2   0   0.2   0.4   0.6   0.8   1
(Times 10E-7)
TIME/SECONDS
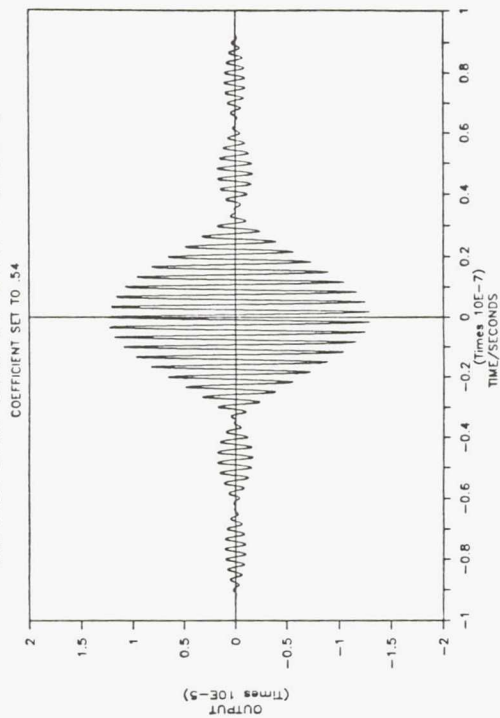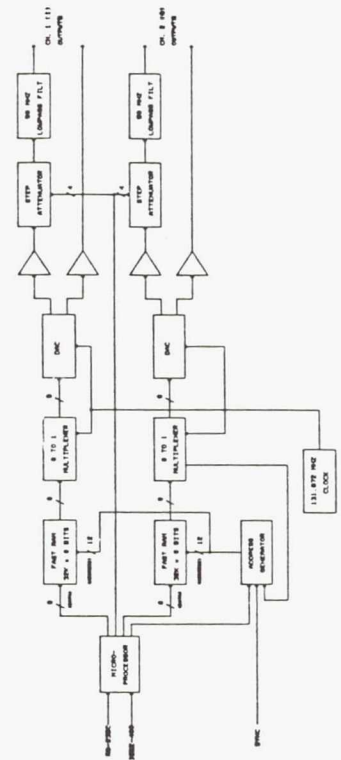
FIGURE 5

FIGURE 6, BLOCK DIAGRAM OF CHIRP WAVEFORM GENERATOR

MULTIPLIER

$S_i(t)$
V_o

$e^{j(W_o t + kt^2)}$ W_o (B, T)

$e^{jW_o(t-\delta)}$

$e^{j(W_1 t - kt^2)}$

CHIRP

$\frac{}{(2B, 2T)}$

$e^{j(W_1 t - kt^2)}$

$e^{j(W_2 t + kt^2)}$ $W_2 = W_o + W_1$ (B, T)

$S_i(t)$

FIGURE 3   CONVOLVE-MULTIPLY-CONVOLVE (C-M-C)
CHIRP TRANSFORM.

IDEAL CMC TRANSFORM OUTPUT

TRANSFORM OUTPUT AMPLITUDE
(Times 10E-5)

2
1.5
1
0.5
0
-0.5
-1
-1.5
-2

-1   -0.8   -0.6   -0.4   -0.2   0   0.2   0.4   0.6   0.8   1
(Times 10E-7)
TIME AXIS (SECONDS)

FIGURE 4

Timebase 20.0 ns/div Delay/Pos 4.90600 us Reference Center Mode Realtime (NORMAL)

FIGURE 10, OUTPUT OF CMC TRANSFORM PROCESSOR
INPUT SIGNAL FREQUENCY= 70.0015 MHz

Timebase 20.0 ns/div Delay/Pos 24.8960 us Reference Center Mode Realtime (NORMAL)
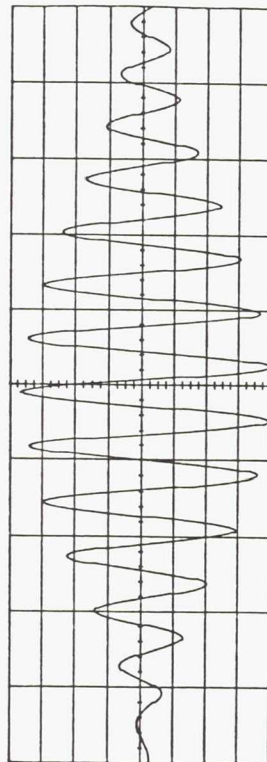
FIGURE 11, OUTPUT SIGNAL WAVEFORM OF CMC PROCESSOR
INPUT FREQUENCY= 75 MHz, OUTPUT FREQUENCY= 70 MHz

Timebase 1.00 us/div Delay/Pos 10.7000 us Reference Left Mode Realtime (NORMAL)
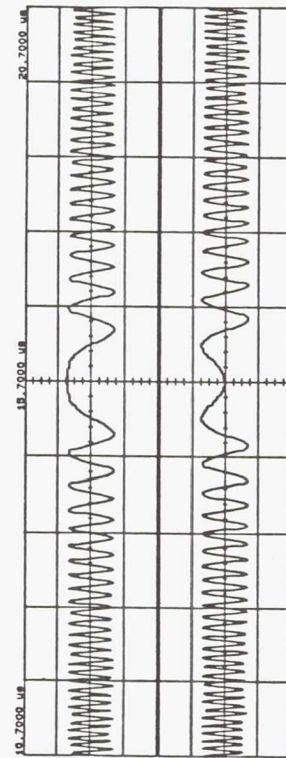
FIGURE 12, I&Q BASEBAND CHIRP WAVEFORMS

FIGURE 7 . TEST CIRCUIT FOR CMC CHIRP TRANSFORM

FIGURE 8, OUTPUT WAVEFORM OF SSB MODULATOR

Timebase 20.0 ns/div Delay/Pos 4.90600 us Reference Center Mode Realtime (NORMAL)

FIGURE 9, OUTPUT WAVEFORM OF CMC TRANSFORM PROCESSOR
INPUT FREQUENCY= 70 MHz

205

Page intentionally left blank

COMSAT LABORATORIES' ON-BOARD BASEBAND SWITCH DEVELOPMENT[*]

B. A. Pontano, W. A. Redman, T. Inukai, R. Razdan, and D. K. Paul
COMSAT Laboratories
Clarksburg, MD   20871-9475

SUMMARY

Work performed at COMSAT Laboratories to develop a prototype on-board baseband switch is summarized. The switch design is modular to accommodate different service types, and the architecture features a high-speed optical ring operating at 1 Gbit/s to route input (up-link) channels to output (down-link) channels. The switch is inherently a packet switch, but can process either circuit-switched or packet-switched traffic. If the traffic arrives at the satellite in a circuit-switched mode, the input processor packetizes it and passes it on to the switch. The main advantage of the packet approach lies in its simplified control structure. Details of the switch architecture and design, and the status of its implementation, are presented.

INTRODUCTION

It is likely that future generations of satellites will incorporate increased use of multiple spot beams. This will lead to increased spacecraft antenna gains, resulting in higher down-link power densities and increased satellite gain-to-noise temperature ratio (G/T). Both of these factors facilitate the use of smaller and lower cost earth stations.

The increased use of multiple spot beams will require additional switching on board the satellite in order to connect up-link beams to down-link beams. This switching can be accomplished either statically or dynamically. When the number of beams is large, static switching becomes impractical and dynamic switching is needed. This can be performed either at RF or IF using a microwave switch matrix (MSM), or at baseband with a baseband switch. The use of MSMs is almost exclusively limited to high-speed time-division multiple access (TDMA) systems, since switching is performed across an entire transponder. Baseband switching is suited for either TDMA or frequency-division multiple access (FDMA) operating in either a continuous or burst mode. Baseband switching may also be coupled with other forms of on-board processing, such as on-board regeneration, decoding and recoding, and

---

[*] This paper is based on work performed at COMSAT Laboratories under the sponsorship of the Communications Satellite Corporation.

demultiplexing and remultiplexing. These additional features can lead to improved performance, greater flexibility, and less complex earth station design.

SWITCH ARCHITECTURE

The baseband switch under development at COMSAT Laboratories is intended to provide switching between six input ports and six output ports. Each port will accommodate bit rates of up to 120 Mbit/s (extendable to 155 Mbit/s). Figure 1 shows the modular structure of the baseband switch, which provides for both interbeam connectivity (input ports switched to output ports) and interservice connectivity (i.e. connectivity between TDMA and FDMA services). This is accomplished within the input and output (service-dependent) modules, as shown in Figure 1. Any mix of FDMA and TDMA inputs or outputs may be accommodated by the baseband switch. Interconnectivity between service types can lead to simplification in the ground segment, since any earth station can employ a single transmission method that is optimized for its own service requirements rather than those of its correspondents.

A number of switch architectures were examined for possible implementation (see Figure 2). These included the conventional space switch and a time switch. It was assumed that either switch would process traffic data on a frame by frame basis (typically 2ms). For the space switch, all input modules transmit simultaneouly; therefore, the physical paths within the switch must be changed to route traffic data from a given input module to each output module. This requires that the switch never simultaneously connect two input modules to the same output module. In contrast, the time switch works by having each input module transmit in a time-division multiplexed (TDM) sequence. Traffic data sent by a given transmitting module are received by all output modules, on either a bus or ring. Only traffic data intended for a given output module are retained by that module. Since the time switch shares (in time) the interconnection medium, the physical connection does not change. For this reason, control of the time switch was considered to be less complex than that of the space switch.

The aggregate information bit rate of the switch is 6 x 120 Mbit/s, or 720 Mbit/s. To allow for TDM guard times and overhead inefficiencies, a switch speed of about 1 Gbit/s (960 Mbit/s) was selected. Mass, power, size, and radio frequency interference (RFI) considerations led to the selection of an optical implementation.

Bus and ring switch topologies were considered for the switch implementation. The selected topology is shown in Figure 3. The traffic data are routed on a high-speed fiber optic ring. The ring topology was selected over the bus topology for traffic routing because the hardware for the input and output modules is less complex. Specifically, the modules in a ring topology process continuous data, while in the bus topology they process burst data. Because of this, data synchronization is less complex for the ring topology, leading to a design with lower power dissipation.

The clock that provides timing is distributed to each module via a fiber optic star network. Bus, ring, and star topologies were examined for this function, and the star topology was found to consume the lowest power considering the combined power of the diode lasers and photodetectors needed for implementation.

The switch operates with a 2-ms TDM frame, as depicted in Figure 3. The configuration processor starts each frame by placing a frame marker and its control information onto the optical ring. Input module 1 is the first module to receive the frame marker and control information. It reads the control information intended for it, appends its traffic data and status information to the configuration processor data, and sends the combined data on to input module 2. Input modules 2 through 6 each read the control data in turn and append their traffic data and status information to the data currently in the frame. Subsequently, each output module reads the traffic data intended for it, appends its status information to the received data, and sends the combined data on to the next module. Output module 6 sends the complete frame to the configuration processor, which reads the status information from each module. The process is repeated for the next TDM frame by the configuration processor placing the frame marker and its control information onto the optical ring.

Figure 3 also shows the static bypass switches which enable a failed module to be taken off line. Redundancy may be provided by using n-for-m modules, with n - m modules in the bypass mode.

Traffic data are routed through the switch using a packet-switched approach. For traffic that reaches the satellite in a circuit-switched manner, the input modules packetize the channels into cells containing eight channels. Each cell contains traffic data for the same output module or modules (for multicast) and is labeled with a destination header in its routing field. The control information from the control processor provides the routing and configuration information to the input modules for cell generation. The routing data provided by the control processor for each input module comprises the routing maps of input channels to output modules on a frame-by-frame basis. In addition, for FDMA transmissions, configuration information is provided that maps up-link carriers to input modules and output modules to down-link carriers. This information changes very infrequently, only when carrier frequency plans are changed.

If the traffic data arrive at the satellite in a packet-switched format, the control processor will only provide framing information, since the routing control is contained within the packet headers. This is the most effective method for operating the switch and results in the least complex on-board hardware; however it does require that the earth stations packetize the traffic data with routing headers.

The format for the packet data that flow in the optical ring is shown in Figure 4. The 2-ms frame is partitioned into 1,667 time slots, each capable of transferring a single cell (data unit) between modules. Each cell contains 9 x 128 data bits. The first 128 bits are the cell header, and the remaining 8 x 128 bits correspond to eight 64-kbit/s traffic channels.

SWITCH DESIGN

The switch, shown in Figure 5, employs a modular design which partitions the core switching functions from the service-dependent functions. The design and implementation effort to date has concentrated on the core switching modules.

Semicustom implementations using standard cell or gate array technologies are being targeted to minimize the size and power requirements of the switch. Each module will require about 20 to 30 chips and will be implemented on a 5 x 5-inch (approximately) circuit board. The core switch module (Figure 6) incorporates two application-specific intergrated circuits (ASICs). Chip 1 is the high-speed switch interface ASIC, and chip 2 is the switch core processor ASIC. This design will permit a module to serve as either an input or output module.

The high-speed switch interface ASIC interfaces at 960 Mbit/s to the optical fiber ring. This is accomplished with one diode laser transmitter on the transmit side for the traffic and status data, and two photodetectors on the receive side, one for data and one for clock. The diode laser used in the design draws 720 mW of power and produces about 1 mW of optical power. The efficiency of these devices is expected to increase significantly in the next few years. This chip accepts the optical clock and optical traffic and control data from the ring and transfers traffic and status data to the ring. To perform these functions, the ASIC establishes and maintains frame synchronization, phase-aligns the clock and data, encodes/decodes packet headers, and interfaces to the switch core processor and data random-access memory (RAM) over the 16-bit traffic data bus.

Because of the speed required, the high-speed switch interface ASIC uses a gallium arsenide (GaAs) technology. It is estimated that this chip will require 6,000 gates and have a total pin count of 84. The estimated power dissipation for the chip is slightly less than 2 W.

The switch core processor ASIC transfers data between the service-dependent module and the core processor module. For circuit-switched traffic, it accepts and executes configuration and routing commands from the configuration processor. The configuration information is sent to the output modules as part of the routing header to identify the appropriate down-link carrier for FDMA transmissions. This chip sorts and routes both the circuit- and packet-switched data to the appropriate output module. For circuit-switched traffic, the switch core processor ASIC in an input module sorts and packetizes the traffic into cells containing eight 64-kbit/s channels to the same destination(s), attaches a routing header, and sends the packet to the RAM for transfer to the high-speed switch interface ASIC. For packet-switched traffic, this step is eliminated. When used in an output module, this chip routes incoming packet data to the appropriate down-link carrier.

The device for the switch core processor will likely be implemented in complementary metal-oxide semiconductor (CMOS) technology having an estimated gate count of 50,000 with a pin count in the range of 256 to 312. The estimated power dissipation for this chip is 2.6 W. Table I gives an estimated breakdown of the power dissipation for the core module.

TABLE I

Core Module	Power Dissipation (W)
HSSI ASIC	1.9
SCP ASIC	2.6
RAM	1.7
Diode Laser (1)	0.7
Photodetector (2 @ 0.6 W)	1.2
Total	8.1

The physical layout of the baseband switch is given in Figure 7. The input and output modules will be housed in two separate sets of trays, each containing one core module and one service-dependent module. This packaging approach permits the input modules to be physically separated from the output modules. Within a spacecraft, for example, it will be possible to locate the input modules close to the receivers and the output modules close to the high-power transmitters. For the distances required, the loss in the fiber optic ring is neglegible. In addition, RFI to or from the optical fiber is nonexistant.

A breadboard circuit (Figure 8) of the critical functions of the switch has been built and tested. The critical functions include the phase adjustment circuit, which aligns the clock and data; integration of the optical to electrical and electrical to optical converters; and fabrication at Gbit/s speeds using surface-mount devices. The phase adjustment breadboard circuitry, using a broadcast clock and ring data, was successfully built and tested using surface-mount devices. The conversion of 1-GBit/s data from electrical to optical and back to electrical was also successfully demonstrated on the breadboard.

CONCLUSIONS

Work continues at COMSAT Laboratories on development of the high-speed on-board baseband switch. The design of the configuration processor module and the high-speed switch interface ASIC is complete. It is expected that the ASIC will be fabricated and tested before the end of 1991. The core processor ASIC design is about 75 percent complete and it is expected that this chip will be fabricated in early 1992. Testing of the core switching functions is expected to begin by mid-1992. The design of the service-dependent functions is expected to be completed by the end of 1992. The complete switch, including the service-dependent modules, will be tested in 1993.
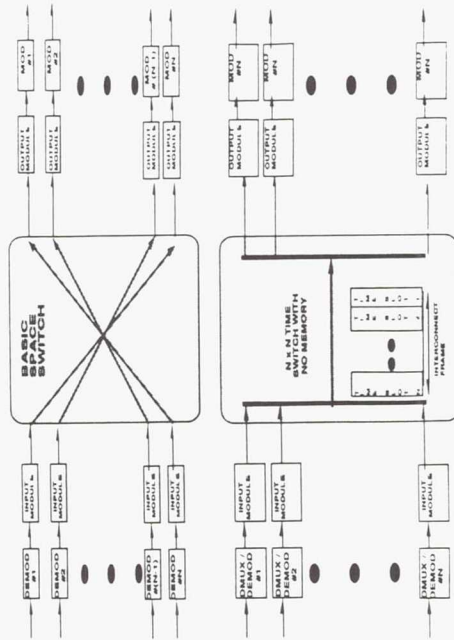
FIGURE 2: EXAMPLE OF TIME & SPACE SWITCH
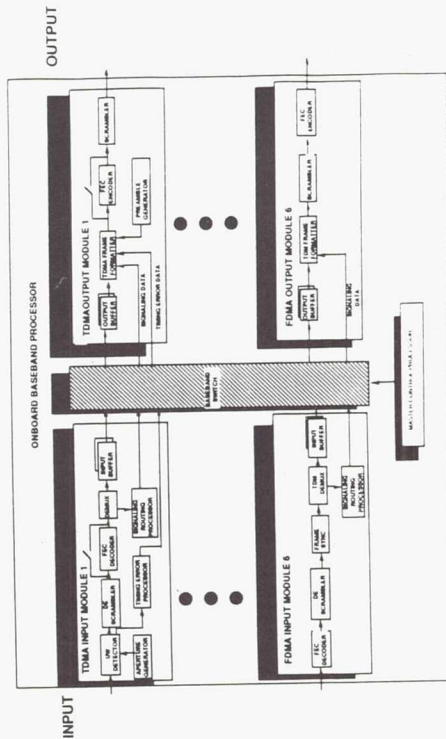
Figure 1: Onboard Baseband Switch

Figure 4: Frame Format

Figure 3: 1-Gbit/s ON-BOARD BASEBAND SWITCH ARCHITECTURE
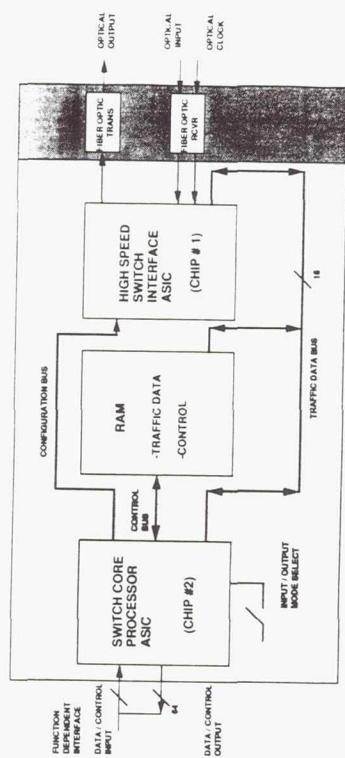
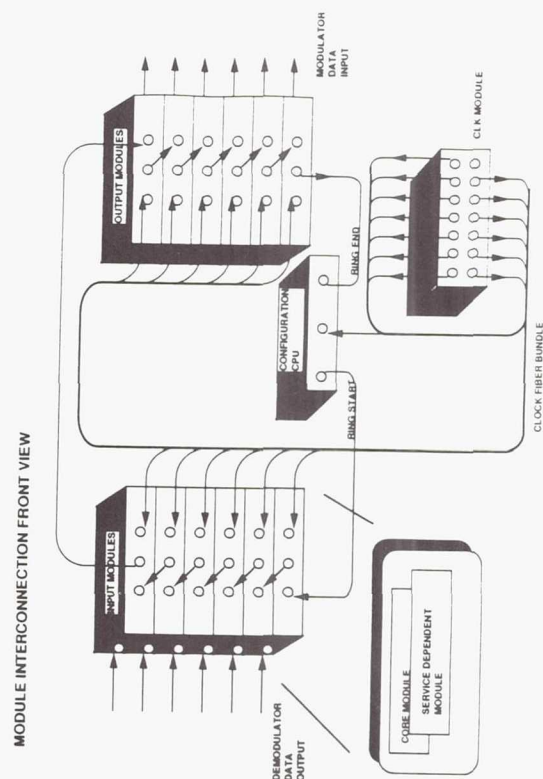FIGURE 6: Core Module Design

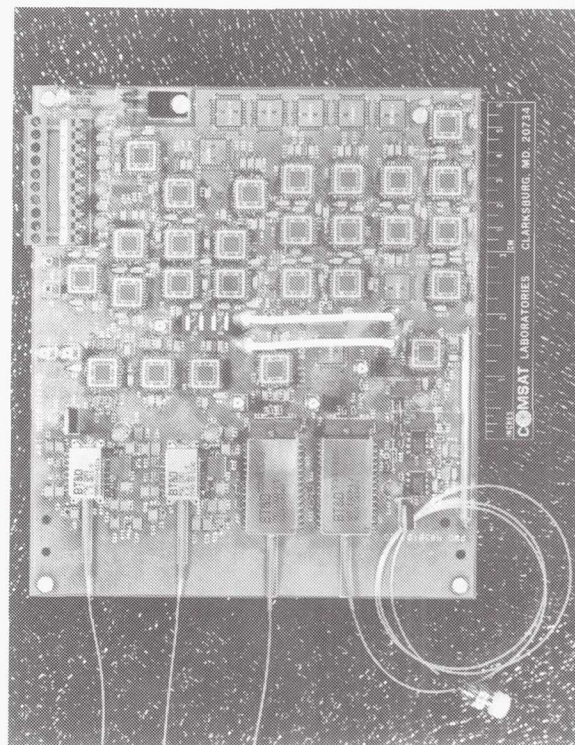Figure 8: Breadboard Circuit

Figure 5: Switch Design

FIGURE 7: Physical Layout of Switch

213

Page intentionally left blank

AN ADVANCED OBP-BASED PAYLOAD OPERATING IN AN ASYNCHRONOUS NETWORK FOR FUTURE DATA RELAY SATELLITES UTILISING CCSDS-STANDARD DATA STRUCTURES

M. Grant

British Aerospace, Space Systems

Argyle Way, Stevenage UK SG1 2AS

A. Vernucci

Space Engineering

via dei Berio 91, Rome Italy 00155

SUMMARY

While preparatory activities for a first-generation European Data Relay Satellite System (DRSS), due for deployment in 1998, are currently being performed, investigations are also taking place to define the technologies and the architectures for advanced DRSSs, featuring the absence of zone-of-exclusion, improved European regional coverage and a high level of interconnectivity and flexibility. The main challenges for such a system are to accomodate a large variety of users, with different requirements in terms of data volumes, bit-rates, service characteristics, etc., and to provide a high degree of flexibility in routing data through the various network links.

Major advances in terms of on-board mass and power savings are expected for digital devices in the next decade, while the same may not occur for RF devices. It is considered appropriate then to exploit the possibilities offered by technology and to propose the use of an On-Board Processor (OBP) aboard each satellite of the DRSS. OBP allows the system designer to individually optimise the various link parameters, to achieve full interconnectivity and flexibility, to accept and process data structures having different multiplexing formats, to terminate useless information (namely "idle frames") on-board and to simplify optical links operation and design.

After introducing a possible DRSS topology and network architecture, the paper discusses an asynchronous network concept, whereby each link (Inter-orbit, Inter-satellite, Feeder) is allowed to operate on its own clock, without causing loss of information, in conjunction with packet data structures, such as those specified by the CCSDS for advanced orbiting systems. The paper then describes a matching OBP payload architecture, highlighting the advantages provided by the OBP-based concept and then giving some indications on the OBP mass / power requirements. This paper is derived from the results of a study performed under a European Space Agency contract.

INTRODUCTION

The European Space Agency (ESA) has in place a number of development programs aimed at establishing an autonomous European manned presence in-orbit. These programs include the development of the Hermes Manned Reusable Space Plane, the Columbus Pressurised Module (an integral part of the International Space Station Freedom), and a Man Tended Free Flying Laboratory. To support the communication requirements of these and other programs, ESA is also developing a Data Relay Satellite System (DRSS) similar to the American TDRS system. The European DRSS is planned to be operational by 1998.

In parallel with this activity, ESA has initiated studies which seek to plot out the strategic future of the In-orbit Communications Infrastructure required to support the space programs of the next century. In this paper, we discuss some of the results of such a study into a future Space Communications Network (SCN), focussing in particular on one
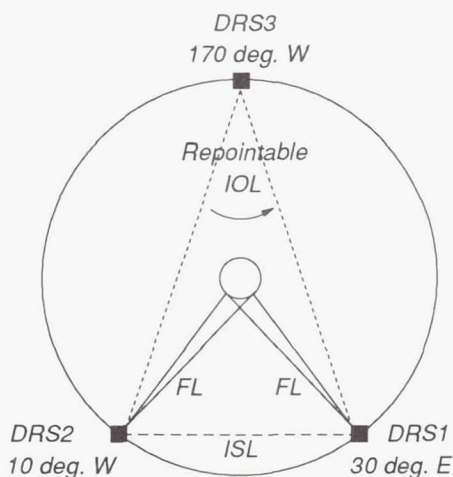
key result, namely the importance of On-Board Processing (OBP) to the success of such a system. The time frame under consideration is 2000 - 2035.

Fig. 1 Present DRSS topology

Before discussing the "near-earth" part of the SCN, we first outline the current DRSS concept. The present DRSS topology is shown in fig. 1. The system consists of two Data Relay Satellites (DRSs) placed at 44 deg. West, and 59 deg. East. Each DRS carries a Feeder Link (FL) at Ka-band (20/30GHz), and Inter-Orbit Link (IOL) accesses at S-band, Ka-band (23/27GHz), and at optical frequencies (0.8μm), the latter being provided on a pre-operational basis. Discussions are ongoing at this time to ensure interoperability with current S-band services on TDRS, and future Ka-band and possible optical services on advanced TDRS. As with the TDRS system, a zone of exclusion exists around the far side of the earth, where communications is not possible with either DRS. This zone extends up to several thousand kilometers altitude for the current DRSS configuration.

AN ADVANCED TOPOLOGY FOR FUTURE DRSSs

In the SCN definition activity, forecasts were produced for the number and type of space programs likely to be undertaken in the timescale under consideration. These forecasts considered every type of space activity from expendable and reusable launchers, through to large space stations and industrial activities. We can summarise the main findings of the study as follows: *i*- the nominal growth scenario projected 19 user satellites of the SCN by 2015, and 33 by 2035; *ii*- each user would require continuous communications with minimal interruption resulting only from link handover; *iii*- standard space link access rates would be defined compatible with bearer services available to users on the ground and the delivery of data should be transparent to the network interworking process; *iv*- the system should seek to deliver data to the ground users in the most cost-effective manner possible.

Fig. 2 The reference near-earth SCN topology

The reference near-earth SCN configuration, selected among a number of topologies traded-off against each other, is shown in fig. 2.

It consists of three DRSs (30 deg. E, 10 deg. W, and 170 deg. W), the first two carrying IOL, Inter-Satellite Link (ISL), and FL payloads and the last one carrying only IOL and ISL terminals. This configuration provides total global coverage for the space users with no zone-of-exclusion.

From DRS3, connection is made via ISL to one of the two DRSs over Europe for data to be downlinked to the ground. Alternative routing paths are feasible with the Pacific DRS having a choice of European DRSs to crosslink to, and each of the European DRSs having the ability to crosslink data onto either FL.

SPACE LINKS TRADE-OFFS

Technologies for the implementation of each of the link types (IOL, ISL, FL) were studied. The following conclusions were arrived at: *i*- the FL should continue to be implemented at Ka-band; *ii*- ISL's would be most efficiently implemented at optical frequencies; *iii*- IOL's would be a mix of optical and S-band technology. The reasons behind this last decision were that S-band is the most appropriate technology to service those space users requiring low data

rate and having omni-directional antennas. All higher data rate services would be best served by a single technology, leading to a homogeneity of IOL terminals, or *accesses*, on the relay spacecraft. Given the rapid advances that are now taking place in optical technology and terminal design and the inherent limitations of RF technologies, it was recommended that the technology used here should be optical. The decision to go optical was re-enforced by the findings that the network would of necessity be regenerative. This requirement arose from the findings on phase noise present in millimetre wave and coherent optical systems. This precludes the transparent carriage of low data rate services; baseband multiplexing of these services is therefore required at the DRS. Given this, there is a very little penalty and great benefits to be gained from making the whole system regenerative.

Probability	99.9%	99.5%	95%	90%
Accesses 2015	17	16	14	13
Accesses 2035	26	25	22	21

Tab. 1 Availability vs. number of accesses

The size of the payload on each DRS is critically dependent on the number of IOL terminals provided. This in turn is a function of the grade of service offered to the users in terms of access availability when a link is requested. Tab. 1 shows the relationship between link availability and number of accesses provided on the 170 deg. W spacecraft. It shows that, for a link availability of 99.9%, 26 IOL terminals are required, on each DRS, for year 2035.

Of course, this availability is realised only if any user can access any terminal and obtain the service he requires. It is not sufficient simply to provide terminals of the same technology. Homogeneity of data rates is also required. To address this issue, three levels of data rates were defined (see tab. 2). The first is the *space link data rate*, defined as the data rate of the service between user and relay spacecraft, including CCSDS packet protocol overheads. The *user data rate* is that rate which is generated from the instrument or unit on board the user spacecraft.

Electrical Rate	Data rate without insert		Space link data rate
	User data rate		
	GI or GII	GIII	
16 Kbps	11.06 Kbps	12.67 Kbps	12.96 Kbps
64 Kbps	44.24 Kbps	50.67 Kbps	51.84 Kbps
1.92 Mbps	1.327 Mbps	1.520 Mbps	1.5552 Mbps
35 Mbps	24.19 Mbps	27.7 Mbps	28.35 Mbps
156 Mbps	107.83 Mbps	123.5 Mbps	126.36 Mbps

Tab. 2 Data rates on the various SCN segments

The *electrical rate* is the rate which results in the ground network, in this case ISDN, after encapsulation of the CCSDS packet. Additionally, it was recognised that many users require more than a single service or channel. Thus services from the user spacecraft are multiplexed onto higher rate bearers. There are many such schemes for defining these bearers. One such scheme provides for three *IOL* bearer rates carrying combinations of the services offered, i.e.: *1.672 Mbps* (1.5552 Mbps + 2*51.84 Kbps + 12.96 Kbps), *30.113 Mbps* (28.35 Mbps + 1.5552 Mbps + 3*51.84 Kbps + 4*12.96 Kbps) and *126.865 Mbps* (126.36 Mbps + 9*51.84 Kbps + 3*12.96 Kbps).

The implications of introducing a scheme such as this is that each terminal is then designed to operate at three standard bearer rates. At the same time no user has to dramatically oversize his terminal package, as the size and mass of an optical IOL terminal is a weak function of data rate. The OBP package can then operate on the received data stream and demultiplex and route services as and when required.

The *ISL* between relay spacecrafts carries a single high data rate multiplex on an optical carrier. Thus, the OBP has complete flexibility on how it orders packets and services. With the user community discussed above, it is conceivable that data rates of up to *1.1Gbps* could be required on ISLs.

The *FLs* are assumed to be multi-channel links, with each channel carrying up to *140Mbps*. This is considered to be the maximum feasible for Ka-band High Power Amplifiers, given the projected development of this technology over the timescale under consideration.

THE ASYNCHRONOUS NETWORK CONCEPT

In the system architecture previously illustrated, the function of the OBP is that of ensuring the proper routing (to an IOL or an ISL or a FL) of the individual communications channels multiplexed over digital streams, operating at different bit-rates. Typically, the need for synchronisation of all such streams would exist, as frame start epochs have all to appear properly aligned at the on-board switch inputs. In previous study activities, it was demonstrated that the synchronisation schemes required for systems based upon multiple satellites interconnected by ISLs are very complex and may also suffer reliability problems.

An alternative approach is that of adopting an *asynchronous* network concept, which is particularly suitable when operating with *packet* communications structures. According to this concept, the on-board switch and the various links independently operate on their own clocks, which are necessarily not identical to one another. Information loss is prevented by the presence of useless or "idle" frames in the data stream. These are transmitted whenever there is no useful information to be delivered, for the sole purpose of maintaining the receiver synchronisation.

In many cases, the presence of idle frames is guaranteed by the natural traffic statistics. Their number can also be increased, should it be required, by overdimensioning the multiplexed stream rate. An overdimensioning will anyway result from the fact that only "standard" bit rates are allowed in the IOLs, as discussed before.

With this scheme, it also becomes possible to dimension the various links taking advantage of the fact that idle frames can be terminated at the OBP, thus increasing the links fill factor, especially in presence of services having a bursty nature.

CONSISTENCY WITH CCSDS-STANDARD STRUCTURES

For an SCN to be implemented in the first decades of the next century, it seems appropriate to propose the adoption of the communications structures specified by the Consultative Committee for Space Data Systems (CCSDS).

Consistent with CCSDS specifications, the communications structures which have been considered are *transfer frames*, each of which comprises one or more *telemetry packets*, which are in close relation with the *source packets* generated by user's instruments. The transfer frames, which are optionally Reed-Solomon encoded, constitute a time-continuous stream, so that *idle frames* are generated when there is no outstanding telemetry packet to be transmitted. Dedicated sequences of transfer frames, not necessarily adjacent to each other, are assigned to carry selected groups of source packets. These are called *virtual channels*.

CCSDS links are typically point-to-point links. In the proposed SCN, which adopts multiplexing techniques on all links, interleaving of information generated by different entities (spacecrafts, ground stations) will necessarily occur. It may be difficult to keep the interleaving pattern under control, it being also dependent on the information statistics. However this is not expected to be a basic problem, because transfer frames are protocol data units independent of each other, in the sense that interleaving of transfer frames of different virtual channels, even if generated by different spacecrafts and/or ground stations, is possible without conflicting with the CCSDS protocol.

Several issues had to be addressed for ensuring the possibility of implementing the desired OBP features and for verifying their consistency with CCSDS specifications. Problems relevant to the availability of *routing information*, to the *segmentation* of the overall link protocol and to the *termination* of idle packets on-board were considered in the study, coming to the conclusion that acceptable solutions can be found to each of them. It was determined that no higher-layer functions (e.g. error correction) shall be provided by the OBP. These will be implemented in the end-to-end protocols. Independent and consecutive frame counts shall be used on each link section. To allow the on-board termination of idle frames, the OBP shall not only operate at transfer frame level, but it shall also examine the content of the transfer frame data field, to interpret the header of the embedded telemetry packets.

As to the transfer frame length, this shall be equal for all frames, to simplify the on-board switch design. This does not pose any constraint, as CCSDS specifications indicate the transfer frame length as a mission set-up parameter. It was considered very appropriate that the telemetry packets length be selected such that an integer number of telemetry packets can be fitted within the transfer frame data field (the CCSDS-specified "synchronous" insertion). In particular, it is proposed that *only one* telemetry packet, comprising a 48-bit header and a 6,960-bit (or 870 octets) data field, be fitted within the transfer frame data field. The resulting transfer frame arrangement is shown in fig. 3.

ATTACHED SYNC MARKER	TRANSFER FRAME				REED-SOLOMON CHECK SYMBOLS
	HEADER	PACKET HEADER	PACKET DATA FIELD	TRAILER	
32	80	48	6960	48	1024

Note: field lengths are expressed in symbols

Fig. 3 Proposed transfer frame structure

THE OBP PAYLOAD

The SCN system supports two logical capacity flows of very different size, i.e. the Forward-Link (FWL) and the Return-Link (RTL) traffic. The inbound and the outbound sections of an ISL will *both* have to support FWL and RTL traffic; therefore a single OBP unit, common to the FWL and the RTL, has been considered. Overall, the OBP can be visualized as an on-board subsystem having the capability of interconnecting several asynchronous input and output streams (IOLs, ISLs, FLs), operating at different bit-rates and comprising CCSDS-standard frames.

OBP ports	IOL	ISL	FL
Input	n @ 13 Kbps 9 @ 1.7 Mbps 4 @ 31 Mbps 13 @ 127 Mbps	2 @ <1.1 Gbps	1 @ 31 Mbps
Output	3 @ 13 Kbps 17 @ 52 Kbps 6 @ 1.7 Mbps	2 @ <1.1 Gbps	5 @ 127 Mbps

Tab. 3 OBP capacity requirements

The OBP shall be capable of terminating idle frames and of routing all other incoming frames to the appropriate output ports. The number of streams to be handled by the OBP is shown in tab. 3 (rounded bit-rate figures are indicated). The number of 13 Kbps IOL streams remains TBD; nevertheless this has a minor impact on the OBP design. The general OBP block diagram, presented in fig. 4, was developed considering that the same design shall be consistent with both the DRS1/2 and DRS3 operational modes. The high-level functions visualised in the block diagram are:

– *Demodulators*, which handle the incoming IF signals (signals travelling over optical links, namely ISLs and the high-rate IOLs, are not to be demodulated, the interface with the OBP being at baseband level).

– *Demultiplexers*, operating on ISLs only. Eight 127 Mbps streams are derived out of each aggregate high-rate (up to 1.1 Gbps) ISL stream (a bit-by-bit multiplexing strategy is adopted to minimise memory requirements). Unique Words are inserted, at multiplexed stream level, to allow the required alignment function.

– *Decoders* (Reed-Solomon), operating on the transfer frames (the attached synchronisation marker is not encoded), utilising the appended check symbols. The Decoders, although placed *in front of* the Synchronisers (next item), have to perform an alignment function, to detect the Start Of Transfer Frame (SOTF flag) within the received stream. Only the transfer frames are to be delivered to the Synchronisers, filling the inter-transfer frame gap (due to the missing synchronisation marker and check bit fields) with "don't care" bits.

– *Synchronisers*, used to align the incoming transfer frames to the unique on-board frame clock (127 Mbps units) and to also perform rate conversion (13 Kbps, 1.7 Mbps and 31 Mbps units). All streams exiting the Synchronisers operate at 127 Mbps. A FIFO-based buffer/alignment device, written with the incoming stream clock and read with the onboard clock, is used to align frames. To counteract the effect of situations where the incoming frame clock rate is higher than the on-board one, it is necessary to terminate some idle frames before they are stored in the FIFO. The Synchronisers shall therefore incorporate logic able to read the header of the telemetry packet embedded in the transfer frame and to accordingly control the FIFO writing operations. They also have to interpret the transfer frame header, thus generating a Transfer Frame Designator (TFD) for each transfer frame. The TFD contains information on both the frame destination-entity (derived from the transfer frame header) and on whether a frame is idle or not (derived from the telemetry packet header).

– *Input Frame Processors*, having the main tasks of generating a Routing Code (RC), subsequently used by the Switching Module to route individual frames to the appropriate output port, and of attaching it in front of each transfer frame, in place of the previously terminated CCSDS synchronisation marker. The RC is a short sequence which unambiguously designates one specific Switching Module output port. The RC is determined on the basis of the TFD pattern and the network route (i.e. direct to a FL or via an ISL) to be followed by each transfer frame to reach the destination entity. This association, decided by the Operations Control Centre (OCC), is stored in a look-up table contained in the On-board Control Unit (OCU), written and periodically updated by the OCC, via a control channel.

– *Switching Module*, which routes the incoming 127 Mbps frames to the appropriate output port. Like switches used for terrestrial ATM applications, it shall be able to provide internal buffering functions, intended to solve conflicts of frames appearing at different inputs but having the same destination port. Each transfer frame has to be handled by the Switching Module preserving the sequencing of frames belonging to the same virtual channel and reducing, as far as possible, the switching delays and their spread.
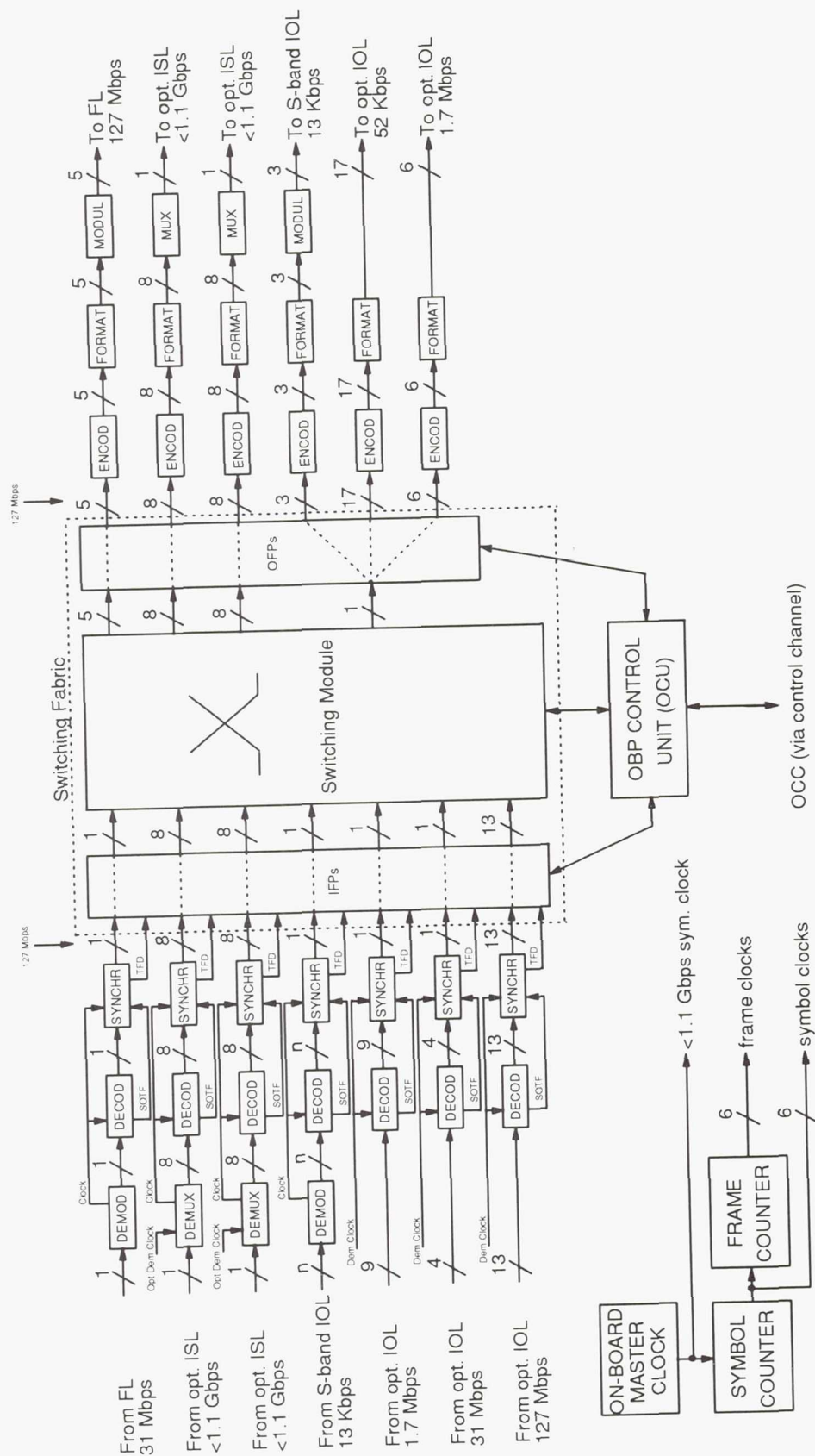
Fig. 4 OBP subsystem functional diagram

- *Output Frame Processors*, intended to change the transfer frame header, to vary the stream bit-rate, to provide buffering where appropriate, and to retime frames at the output of the Switching Module.

- Transmissive equipment, i.e. *Encoders*, performing Reed-Solomon encoding of transfer frames and appending the check bit field, *Output Frame Formatters*, having the simple task to attach the synchronisation marker to the encoded frame, *Multiplexers* (bit-by-bit), only for ISLs, and *Modulators*, for the low-rate IOLs and the FL.

- *OBP Control Unit*, having the main task of keeping look-up tables containing frame routing information. In particular, it controls the associations between the destination entity and the OBP output port. The OCU operates under instructions of the OCC, via a control channel. Look-up tables updating may be expected at all times when a new connection is set-up or when it has to be rerouted for inter-satellite visibility problems.

OBP IMPACT ON PAYLOAD IMPLEMENTATION

From the analyses performed, it was possible to preliminarily evaluate the main OBP parameters, as shown in tab. 4.

Year	Technology	Power (W)	Weight (Kg)	Structure
1990	commercial	185	74	assembly
2015	qualified	4	1.5	board
2035	qualified	0.12	n.a.	chip

Tab. 4 OBP capacity requirements

The most interesting result is the effect of components integration. With today technology, the OBP assumes the aspect of an assembly; however it is expected that, via a massive utilisation of ASIC devices, the OBP can become a simple board in 2015 or even a single chip in 2035. The OBP impact on the overall payload becomes then virtually negligible.

CONCLUSIONS

This paper considers the applicability of OBP techniques to an advanced DRSS, constituting the near-earth part of the SCN. OBP has been found to be beneficial with regard to the:

- *Overall network performance*. The payloads utilised in an SCN typically have to support a wide variety of links characterized by different optimal link parameters (bit-rate, multiplexing, information bundling, modulation, coding, etc.) and heavy interconnectivity and flexibility requirements. OBP can accomodate all such requirements, allowing, in addition, to terminate useless information on-board.

- *Network synchronisation*. A formidable problem is represented by the requirement for synchronising the various streams. This problem can be overcome by adopting packet structures (e.g. the CCSDS standard) in conjunction with OBP, allowing the implementation of an SCN where all links have no mutual synchronization requirement.

- *Optical links*. An SCN will largely rely upon optical links, both for the IOLs and for the ISLs. Due to the difficulty of implementing coherent optical links, with an analog interface, baseband operation appears to be a must, particularly for the interface between the ISL and the other payload sub-systems.

The CCSDS standard has been reviewed and it has been found that, by appropriate selection of parameters, there are no basic inconsistencies between it and the OBP operational mechanisms.

An OBP payload operating in an asynchronous network concept has been proposed; it utilises a Switching Fabric realised with the same techniques which will be used for the implementation of terrestrial broadband networks (B-ISDN). After a review of present technology and expected developments, the OBP has been tentatively sized masswise and powerwise. It was concluded that, by the next century, OBP will be implementable with neglectable impacts on the payload mass and power, while providing important benefits with regard to overall system efficiency.

ACKNOWLEDGMENTS

221

Page intentionally left blank

ON-BOARD PROCESSING FOR TELECOMMUNICATIONS SATELLITES

P.P. Nuspl and G. Dong
INTELSAT
WASHINGTON, D.C., 20008-3098

SUMMARY

In this decade, communications satellite systems will probably face dramatic challenges from alternative transmission means. To balance and overcome such competition, and to prepare for new requirements, INTELSAT has developed several on-board processing techniques, including Satellite-Switched TDMA (SS-TDMA), Satellite-Switched FDMA (SS-FDMA), several Modulators / Demodulators (Modem), a Multicarrier Multiplexer and Demodulator (MCDD), an IBS / IDR BaseBand Processor (BBP), etc. Some proof-of-concept hardware and software were developed, and tested recently in the INTELSAT Technical Laboratories. This paper presents these techniques and shows some test results.

INTRODUCTION

Communications satellites will probably face dramatic competition with alternative (terrestrial/undersea) transmission means. More sophisticated and flexible user-oriented satellite system architectures are being studied and developed to minimize the overall system cost (space and ground segment), and to meet the requirements for low-cost, smaller earth terminals to directly access satellites at low-to-moderate data rates. Reconfigurability and adaptability to different traffic scenarios are also important. Shorter terrestrial tails are needed in many services. Such requirements result in putting as many features as possible into the satellite payloads, and implementing some suitable access / modulation / coding schemes to improve link budgets.

On-Board Processing (OBP) systems, in the form of significant conditioning of traffic signals, are appropriate solutions to these problems, since OBP increases the flexibility of resource utilization and improves link performance [Ref. 1]. OBP offers alternatives to the approach of merely increasing the transmitted power and G/T of receivers. Under its R&D Programs, INTELSAT has developed several on-board processing techniques, including SS-TDMA (operational, and an advanced form), SS-FDMA, Modems, MCDD, IBS/IDR BBP and other items. [IBS is the International Business Service and IDR stands for Intermediate Data Rate, INTELSAT's range of public switched telephony services.] SS-TDMA and IBS/IDR BBP improve the connectivity and flexibility of these services; the Modems and MCDD improve the link performance. Some of these are good for both performance and connectivity improvements. There are some which are simpler and low-risk technologies, and can be specified in the near future. For others, there are some technical problems and system issues which should be resolved before the technologies can be used on-board. Some of

these are suitable for use with transparent transponders but most are in the regenerative class.

Proof-of-concept hardware and software were developed under contracts, and tested recently in the INTELSAT Technical Laboratories. This paper describes these techniques and shows some test results.

OBP FOR TRANSPARENT PAYLOADS

OBP with transparent payloads mainly includes SS-TDMA, where 4-GHz signals with 72-MHz bandwidths are routed from beam to beam, and SS-FDMA, where smaller channels are formed, routed and reformatted. Since regeneration is avoided, these systems are simple and have less risk for on-board application in the near future. The benefits are better performance and higher connectivity.

Satellite-Switched TDMA

In Satellite-Switched TDMA (SS-TDMA), the uplink signals from the satellite receiving beams are demultiplexed in typically 80-MHz bands at RF and sent to the SS-TDMA Switch Matrix which maps the input signals to output beams dynamically. Since the HPA handles only one signal, capacity can be used more efficiently than in multicarrier FDMA.

The Microwave Switch Matrix (MSM) of the INTELSAT VI communication subsystem is an example of SS-TDMA applications [Ref. 2]. This MSM is capable of routing individual bursts of traffic between various satellite beam inputs and outputs. There are actually three functional units associated with the operation of the MSM: the microwave switches, the distribution control unit (DCU), and the timing source. The basic interconnections among the three units are shown in Figure 1.

The MSM payload on INTELSAT VI satellites is a solid state unit which takes advantage of MIC fabrication technology. This MSM has 10 input lines and 6 output lines. The 10 input lines are preceded by a ring-redundancy network which is made up of coaxial R switches. With central symmetry and flexible routing of the input ports through the network, there is minimal signal leakage. Only 6 of the 10 input lines to the matrix are active at any time. The matrix uses a bi-planar coupled crossbar configuration to achieve maximum interconnectivity. The MSM Switching Junction is shown in Figure 2. The input and the output planes are connected through quarter-wave 10-dB directional couplers and PIN diode attenuators. The directional couplers, while increasing the matrix insertion loss, reduce the VSWR and provide good isolation between interconnection points. The PIN diode attenuators act as the RF switching elements. The input preamplifiers, with 16 dB gain, are used to overcome the additional insertion loss caused by the directional couplers. The DCU provides dynamic controls for the MSM. Switch configuration information, called burst time plans, can be up-linked and stored in three DCU memories. The Timing Source provides all the timing signals for the DCU and the switch matrix.

An Advanced Satellite Switching Center has also been developed under R&D contract with NEC (Japan), some years ago [Ref. 3]. It has the same three functional units: MSM array, DCU and Timing Source. A major improvement consists of a redundant design with two 6 x 4 planes passively combined, and replicated again, to achieve an 8 x 8 matrix which has no single-point failure mode. The most significant innovation is the use of dual-gate FET switch modules, which in fact provide very stable and consistent gains (instead of losses). The MSM topology uses directional couplers with optimal ratios to achieve the
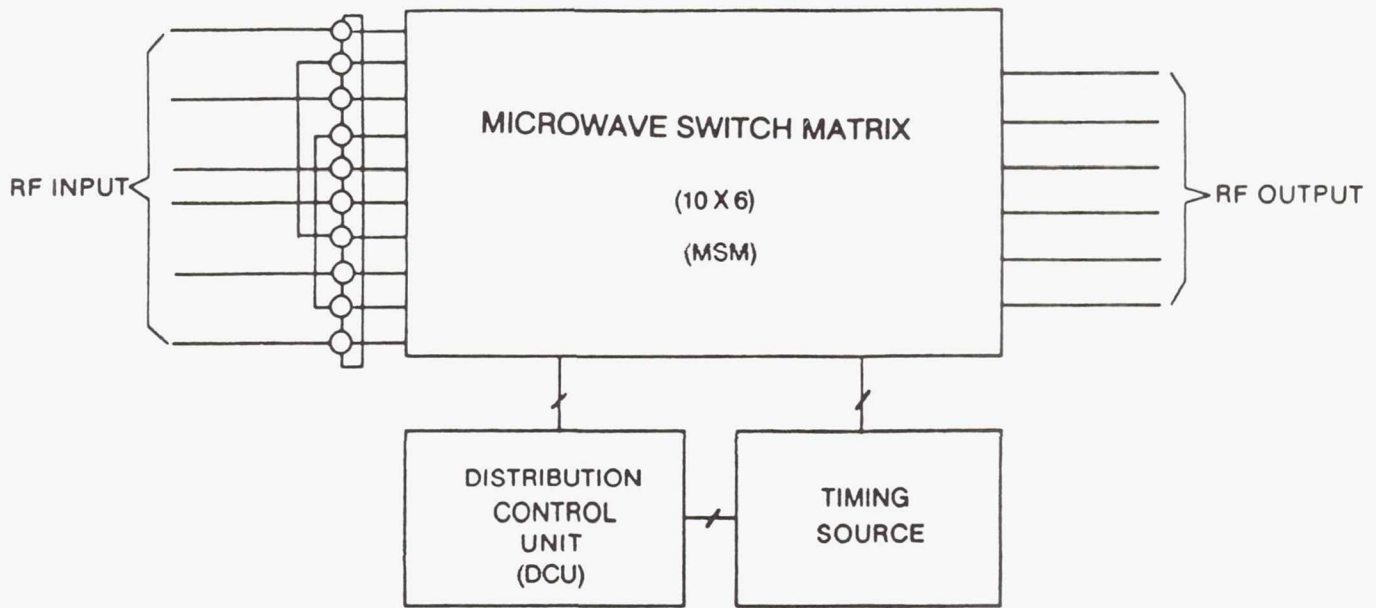
Figure 1 Interconnections among MSM, DCU and Timing Source
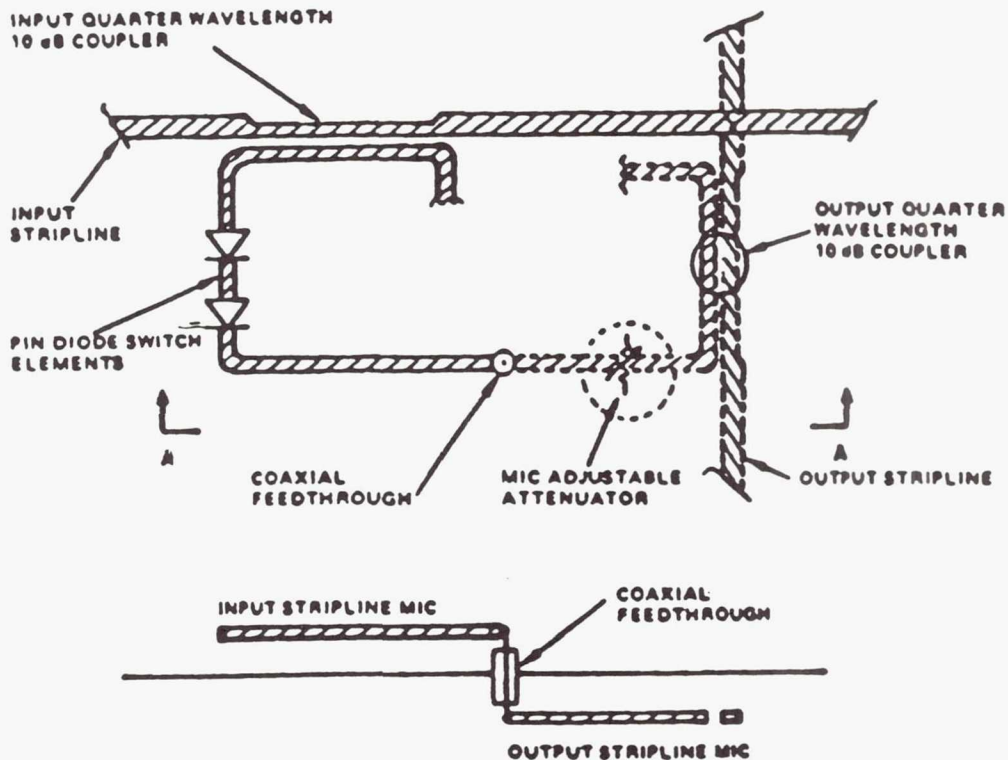

Figure 2 MSM Switching Function Detail

lowest insertion loss.

In a more recent advancement, on-board non-interfering diagnostics have been added, so that appropriate operations and status of the array can be made known to the ground control

225

station. Details have been reported previously [Ref. 4].

Satellite-Switched FDMA

SS-FDMA is an alternative means in the frequency domain to enhance connectivity for FDMA services. In the existing transponder structure, the input demultiplexing and output multiplexing are performed on the transponder channels, i.e., the transponder bandwidths are taken as the elementary minimum bandwidths. With the SS-FDMA concept, the signals in a transponder bandwidth are further demultiplexed into a number of narrower subbands, some of which are multiplexed before the on-board HPA's in a "one-HPA-for-many-duplexed-channels" scheme.

An SS-FDMA package consists of demultiplexers, a static switch array, and multiplexers. A demultiplexer can be a bank of filters with variable bandwiths and variable center frequencies (VBVCF). A switch array consists of the single-pole-multiple-throw (SPMT) switches and crossbar switch matrices; it maps the inputs from the demultiplexer to its outputs. To have a reasonable number of HPAs, a frequency multiplexer combines several subchannels.

Under R&D contract to KDD (Japan), INTELSAT has developed an SS-FDMA technology demonstrator in the form of a VBVCF diplexer (2-channel multiplexer) which utilizes lithium tantalate (LiTaO$_3$) SAW filter technologies and Gallium Arsenide (GaAs) switch technologies [Ref. 5]. Figure 3 shows the layout of this unit. Its main parameters and characteristics are listed in Table 1.

With SS-FDMA processing, the connectivity between the input and output ports is achieved for the narrower subbands in the FDMA services. With this feature, the uplink FDMA signals of mixed high- and low-power
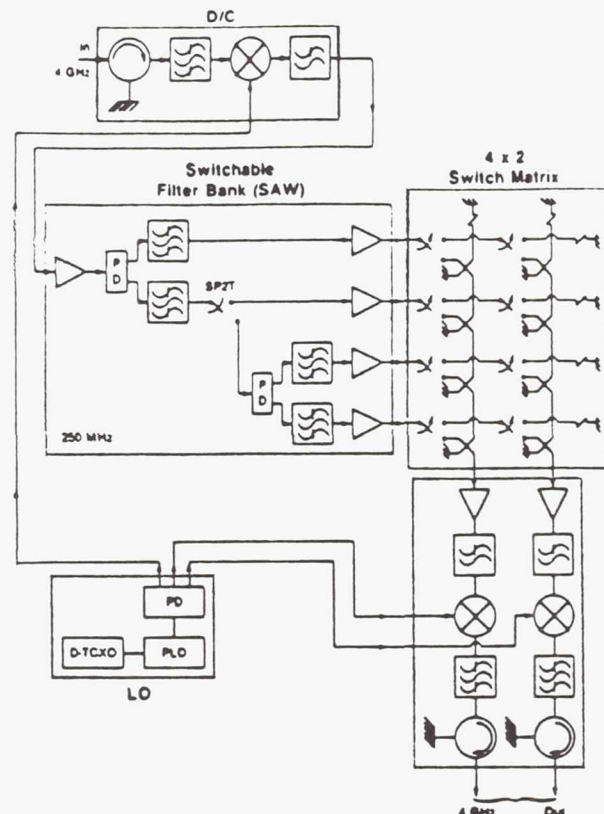
Figure 3 INTELSAT SS-FDMA Demonstrator

densities can be separated before going into the HPA. The performance for the low-power density carriers is improved and the HPA power is used efficiently, with less backoff.

OBP FOR REGENERATIVE PAYLOADS

On-Board Processing for regenerative payloads includes demodulation, baseband switching, and remodulation. On-board demodulation / modulation improve the link performance and isolate the downlink from the uplink. Baseband switching provides better connectivity and a high degree of flexibility. Demodulation for the IDR and TDMA services and the MCDD for IBS and IDR services are discussed below.

In transparent transponders, the uplink noise and interference in the receiver is

226

amplified and retransmitted in the downlink. At the E/S receiver the total noise is the summation of the downlink noise and interference, and the retransmitted uplink noise. The system BER performance is determined by the total $E_b/N_o)_t$.

$$\frac{1}{\dfrac{E_b}{N_o}\Big|_t} = \frac{1}{\dfrac{E_b}{N_o}\Big)_{up}} + \frac{1}{\dfrac{E_b}{N_o}\Big)_{dn}} \qquad Eq. \ 1$$

For the given BER performance, say 1×10^{-6}, the relation between the uplink $E_b/N_o)_{up}$, and downlink $E_b/N_o)_{dn}$ is shown in Figure 4. The curve without regeneration varies slowly with both $E_b/N_o)_{up}$ and $E_b/N_o)_{dn}$, and for a large range the total noise is sensitive to both uplink noise and downlink noise, in this non-regenerative case.

In regenerative systems, the uplink data is demodulated on-board the satellite and modulated on the downlink carrier. From the standpoint of BER performance analysis, all regenerative links may be regarded as a pair of Binary-Symmetric-Channels (BSCs) in cascade. Figure 5 (A) illustrates this schematically, where P_{bu} and P_{bd} denote the information-BERs of the uplink and downlink BSCs, respectively. It is relatively easy to show that the cascaded BSCs reduce to an equivalent BSC (see Figure 5 (B)) whose information-BER P_b is given by:

$$P_b = (1-P_{bu})\,P_{bd} + P_{bu}(1-P_{bd}) \qquad Eq. \ 2$$
$$= P_{bu} + P_{bd} - 2P_{bu}P_{bd}$$

For most practical cases, both P_{bu} and P_{bd} are much less than 1, and $2P_{bu}P_{bd}$ is much less than P_{bu} or P_{bd}, and the above equation is well approximated by Equation 3:

$$P_b = P_{bu} + P_{bd} \qquad Eq. \ 3$$

The total BER is a summation of uplink BER and the downlink BER. For the given BER of 1×10^{-6}, the relation between the $E_b/N_o)_{up}$ and $E_b/N_o)_{dn}$ is also shown with regeneration in Figure 4 for comparison. It is clear from the figure that the curve is very steep with changes of $E_b/N_o)_{up}$ and $E_b/N_o)_{dn}$. The curve is mainly determined by $E_b/N_o)_{up}$ (uplink limited) or $E_b/N_o)_{dn}$ (downlink limited). Only in a very small range is the system BER determined by $E_b/N_o)_{up}$ and $E_b/N_o)_{dn}$ together, but this is the best operating range. OBP allows operation with reduced $E_b/N_o)_{dn}$ and much lower $E_b/N_o)_{up}$.

TDMA Modem

Under an R&D contract to MELCO (Japan), INTELSAT has developed an On-Board Modem for TDMA operation at 120 Mbit/s [Ref. 6], and it was tested recently in the INTELSAT Technical Labs. The On-Board Modem contains a Demodulator and Modulator for burst-mode QPSK 60 Msymbol/s signals.

The Demodulator diagram is shown in Figure 6 (A). The demodulation circuit is a coherent detector. The carrier recovery circuit consists of a times-four multiplier, a tank-limiter with AFC (Automatic Frequency Control) and a divided-by-four circuit. The symbol-timing-recovery (STR) circuit consists of the IF squaring circuit and tank-limiters. The burst 3950 MHz RF signal is fed to the RF channel which includes the downconverter, the IF roll-off filter and the AGC circuit. The RF chain converts the RF signal frequency from 3950 MHz to an IF of 141.1 MHz. The output IF signal is divided into two parts; one is fed to the demodulation circuit and the other is fed to the multiplier in the STR circuit. The frequency multiplier generates times-four and times-two signals, and parallel filters extract the components of the carrier and symbol timing

227

Table 1 Performance of Diplexer

PARAMETERS		CHARACTERISTICS
1-dB Bandwidth	B1	≥ 35 MHz
Center Frequencies	f1	232 MHz
	f2	268 MHz
Guard Band	BG1	≤ 1 MHz
Transition Bandwidth (1 to 30 dB)	F1	≤ 1 MHz
1-dB Bandwidth	B2	≥ 17 MHz
Center Frequencies	f3	259 MHz
	f4	277 MHz
Guard Band	BG2	≤ 1 MHz
Transition Bandwidth (1 to 30 dB)	F2	≤ 1 MHz
Minimum Insertion Loss		1.5 dB
Insertion Loss Variation (Channel to Channel)		≤ 0.7 dB
Out-of-Band Attenuation		> 30 dB
Group Delay Variation		< 10% of Minimum Group Delay
Amplitude Ripple		1 dB p-p
Phase Ripple		≤ 6 ° p-p
Input VSWR		≤ 1.3
Output VSWR		≤ 1.3

signals. The frequency variation of the IF signal is tracked by an AFC circuit in the carrier recovery circuit. The recovered carrier is used as a reference frequency signal for coherent detection in the demodulator. Demodulated P and Q signals go to the regeneration circuit where each of them is converted to a digital signal.

The Modulator is shown in Figure 6 (B). The P and Q streams and their clock of 60.416 MHz (data rate is 120.832 Mbit/s) are received by the retiming circuit. P and Q streams are synchronized by the Clock. The P and Q bit streams from the retiming circuit go to the QPSK modulator. The carrier signal of 141.1 MHz comes from the Test Set through a switch and bandpass filter. This switch is
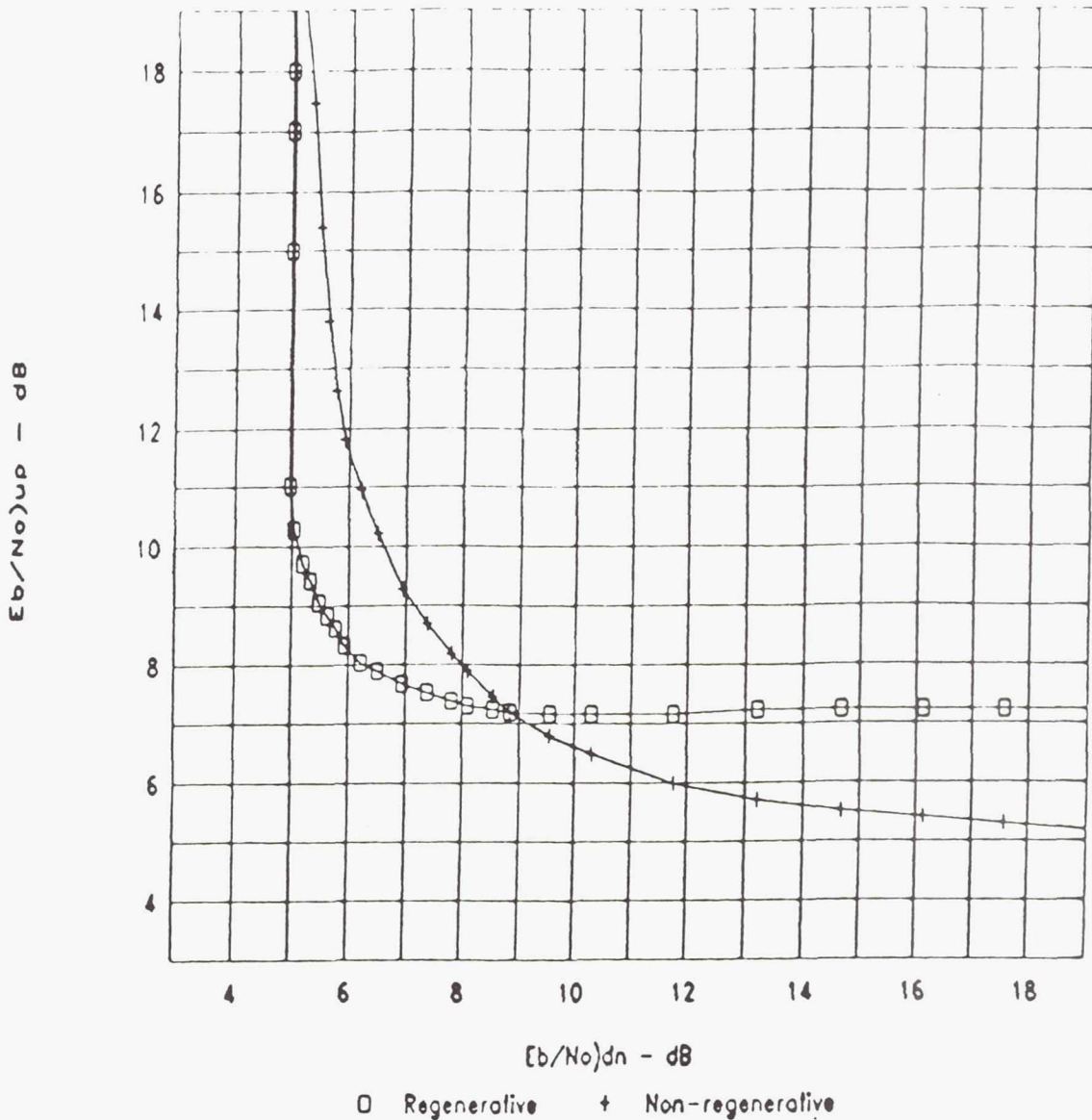
Figure 4   Uplink to Downlink Relationship
(No OB Modem, BER = 10^{-6}, calculated)
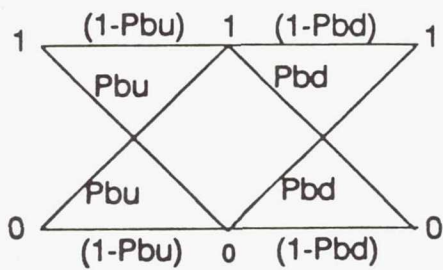
Figure 5 (A)   Uplink BSC and Downlink BSC
in Cascade
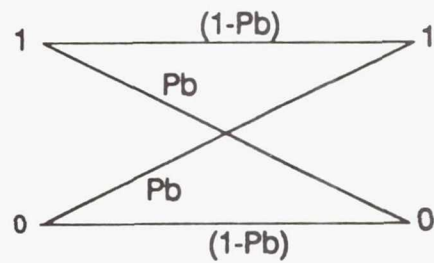
Figure 5 (B)   Equivalent BSC
of Regenerative Link
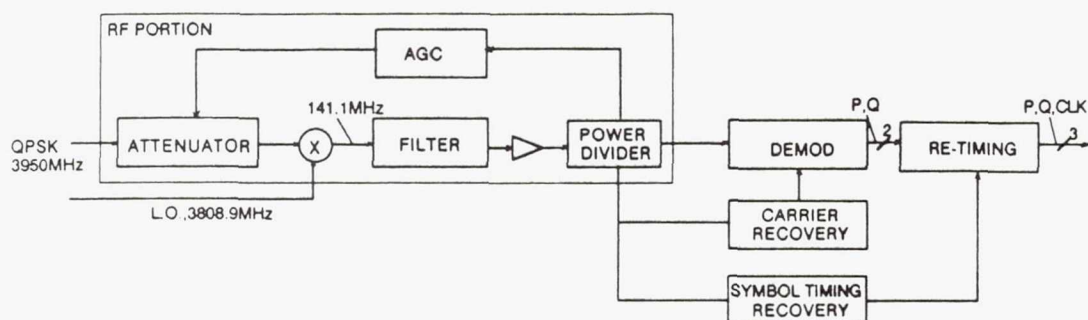
Figure 6 (A) On-Board Demodulator
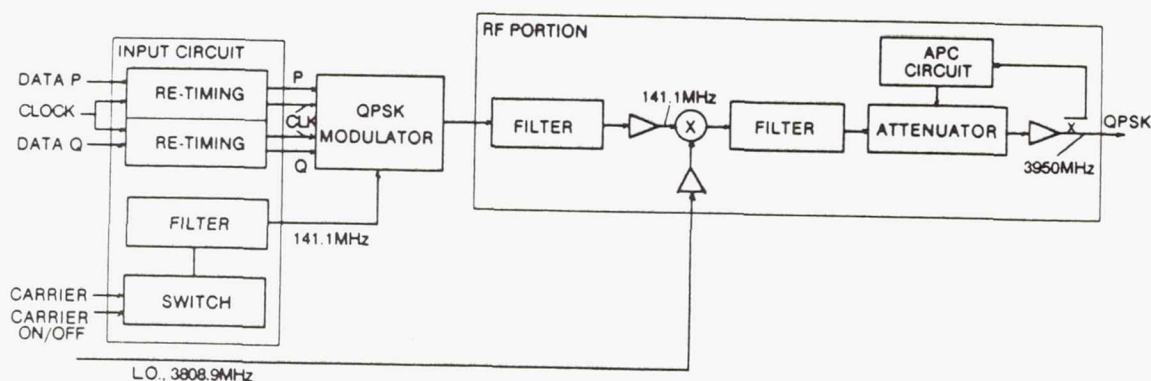
Figure 6 (B) On-Board Modulator

controlled by the carrier on-off signal from the Test Set and it controls the output of the modulator. The modulated signal from the QPSK modulator passes through the IF filter, amplifier and upconverter in which it is converted into an RF signal at 3950 MHz. The LO signal of the Converter which comes from the Test Set is 3808.9 MHz. The RF signal is filtered, amplified and output.

The main items of the TDMA Modem performance were also tested recently at INTELSAT H.Q. and included: Bit Error Ratio (BER) versus E_b/N_o, BER versus carrier frequency offset, BER versus Local Oscillator (LO) frequency offset, BER versus clock frequency offset, BER versus input signal level variation, carrier phase and amplitude

variations and on/off isolation. The E_b/N_o relationship between uplink and downlink for the On-Board Modem for the given BER was also determined.

Figure 7 shows the uplink BER versus E_b/N_o curves of the test results. The specification BER curve is also printed in the figure for performance comparison. The uplink BER performance indicates mainly the On-Board Demodulator performance. The BER versus E_b/N_o results in burst mode and continuous mode are similar. Compared to its specifications the E_b/N_o needs to increase 2.1 to 4.7 dB. The degradation reduces to the range 0.7 to 1.6 dB when the LO frequency offset is at about -125 kHz or the carrier frequency offset is + 125 kHz.

230
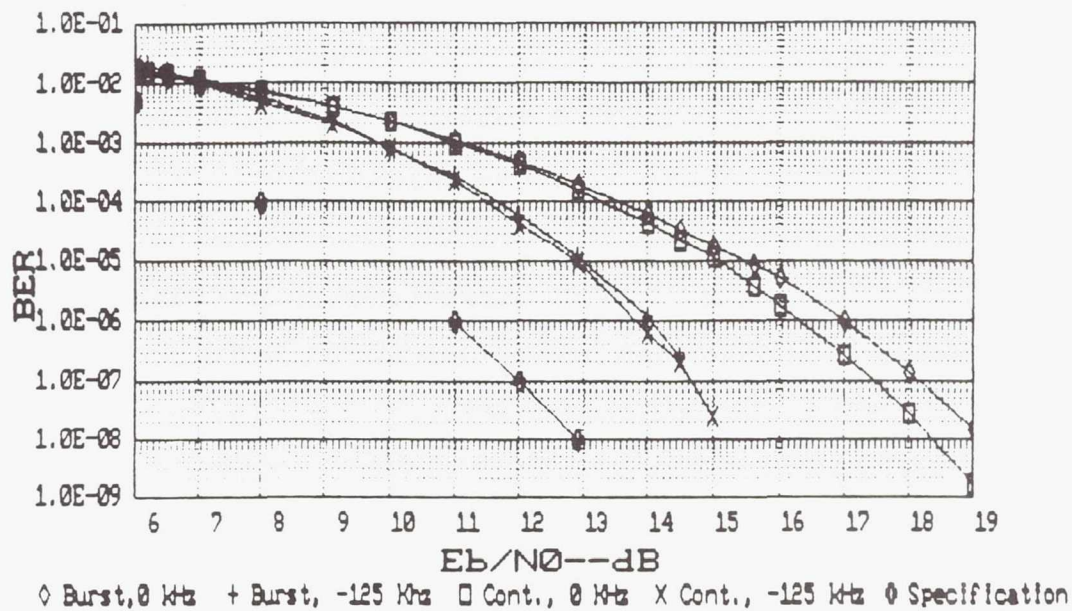
Figure 7 Uplink MODEM BER vs E_b/N_o
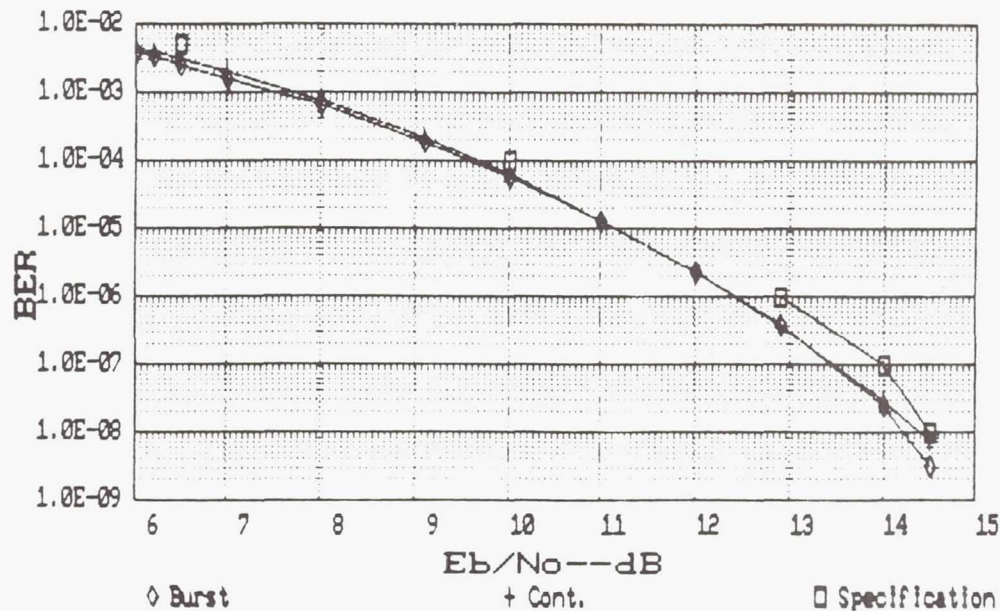
Figure 8 Downlink MODEM BER vs E_b/N_o

Figure 8 shows the downlink BER versus E_b/N_o curves, and the specification BER curves are also shown for comparison. The downlink BER indicates mainly the On-Board Modulator and the E/S Demodulator performance. The BER versus E_b/N_o for burst-mode and continuous mode are very similar. The BER versus E_b/N_o as measured is better than the specifications, for the burst mode about 0.3 to 0.6 dB better than the On-Board Modem specification.

The relationship between the uplink and downlink E_b/N_o was tested for the given BER and the results are shown in Figure 9. Although there is degradation, the relationship

231
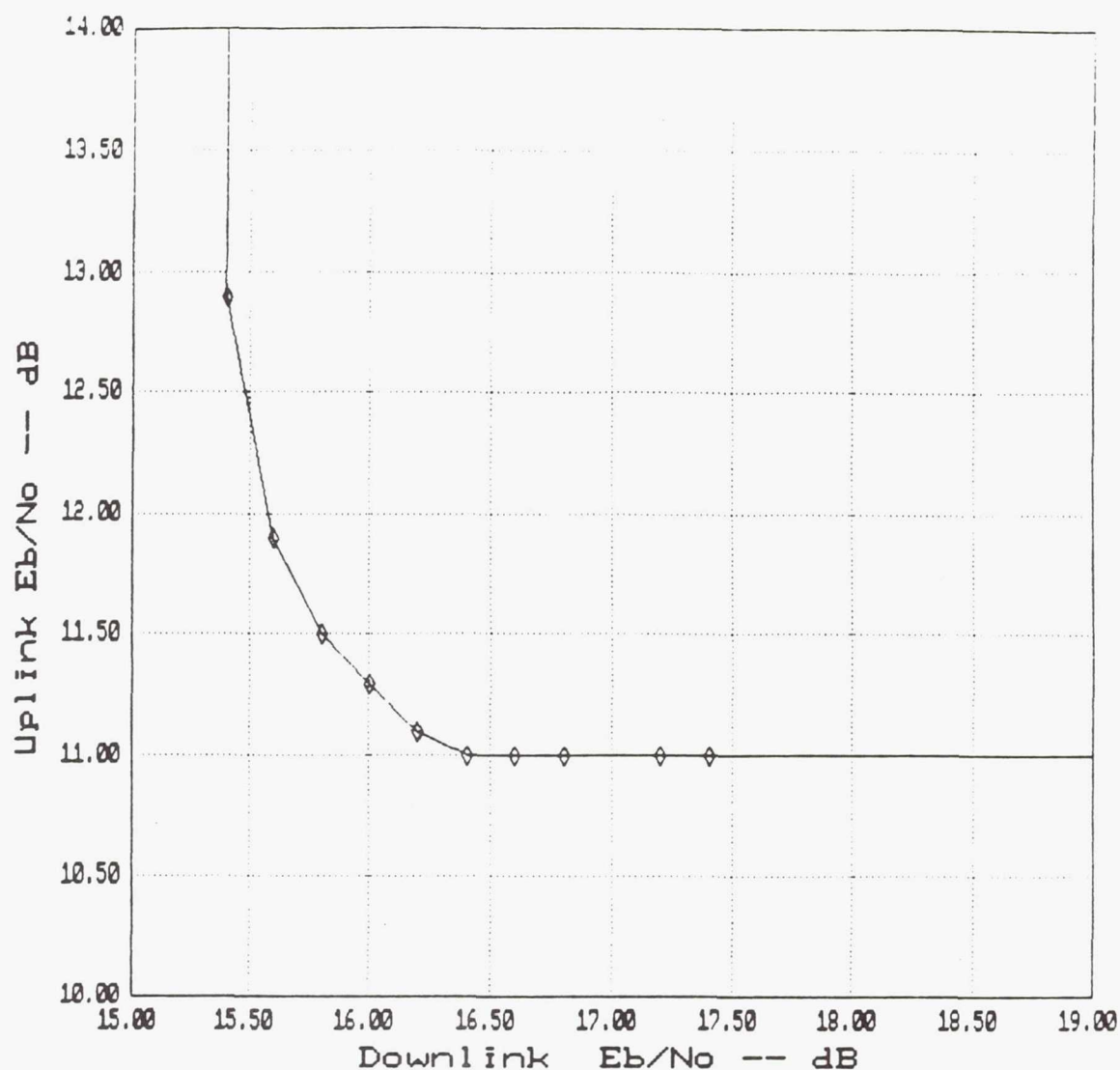
Figure 9 Uplink and Downlink Relationships
(with OB Modem, BER = 10^{-5}, measured)

is very close to that in Figure 4, i.e., the uplink and downlink are well isolated from each other.

MultiCarrier Demultiplexer and Demodulator

Under R&D contract to TELESPAZIO / ALCATEL (Italy/France), INTELSAT has sponsored a proof-of-concept unit of a MultiCarrier Demultiplexer Demodulator (MCDD) [Ref. 7]. The general structure of a MCDD consists of two main blocks: the demultiplexer and the demodulator. The demultiplexer separates the channels and down-converts them to baseband. The demodulator is a single-channel demodulator that recovers the transmitted bit stream and outputs it to a baseband switch matrix. The bit rate of this MCDD can not be varied and only one channel

232
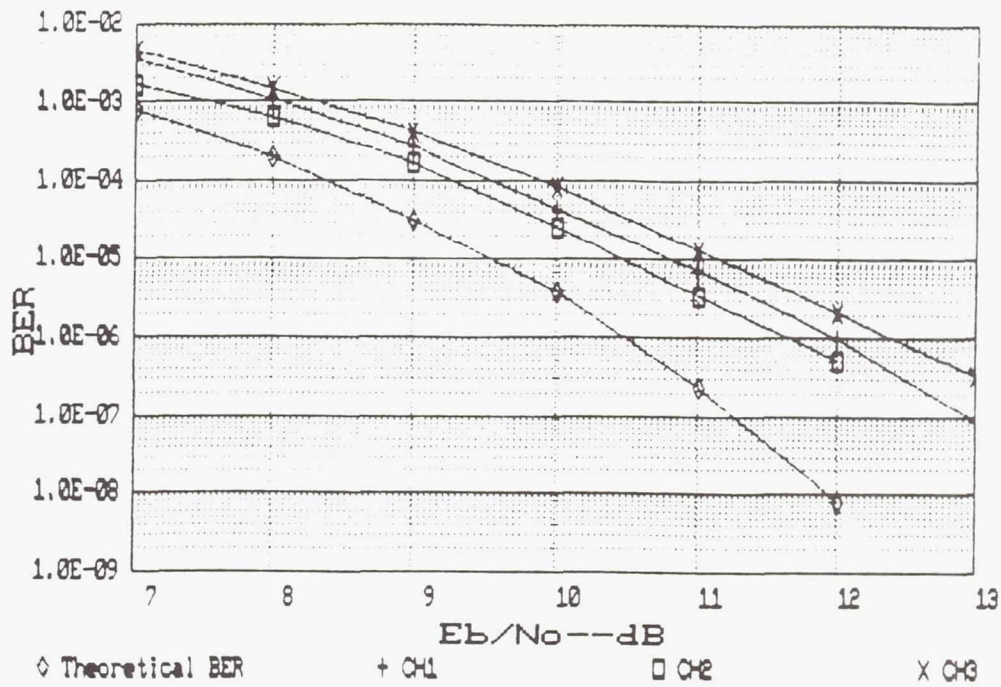
Figure 10 MCDD BER vs E_b/N_o
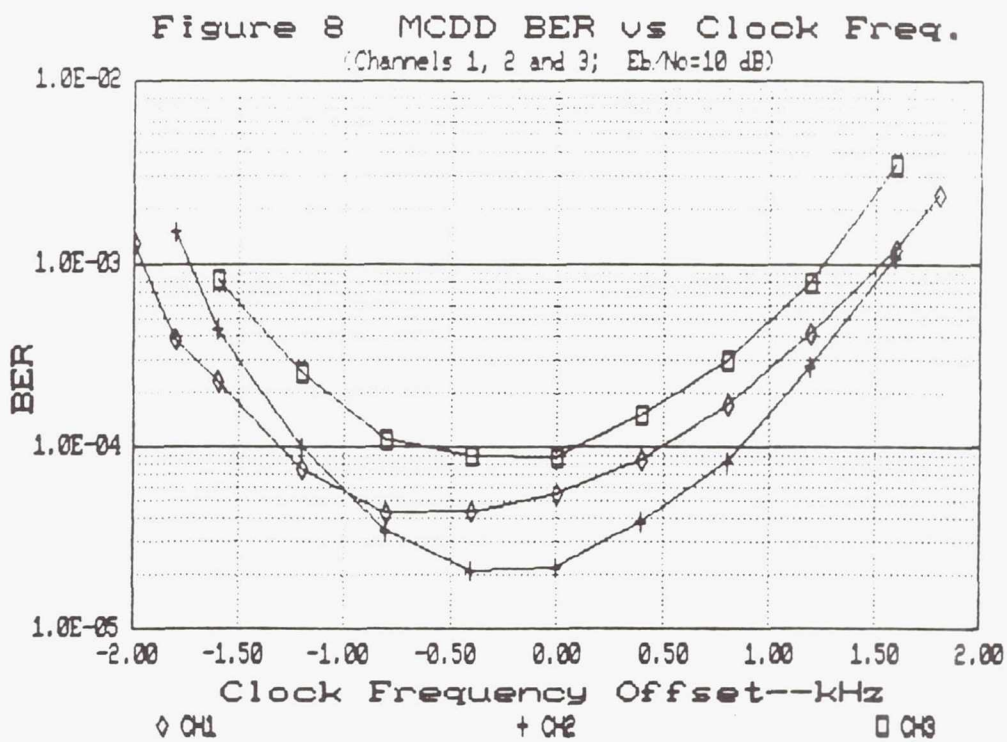
Figure 11 MCDD BER vs Clock Frequency Offset
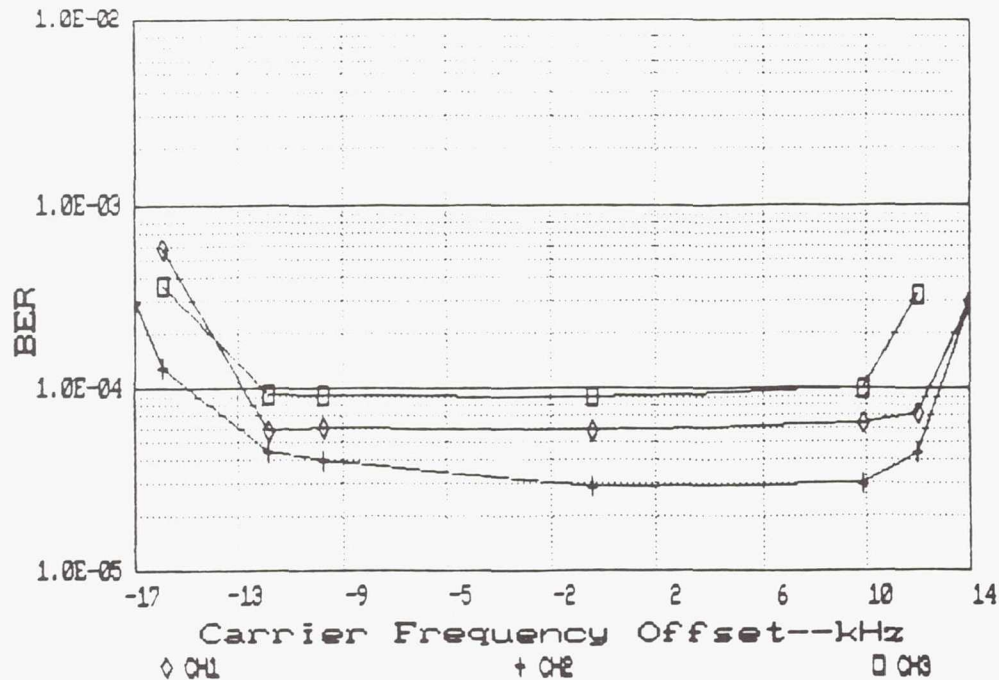
Figure 12 MCDD BER vs Carrier Frequency Offset

can be processed at any one time. The FDMA signal with a 10-MHz bandwidth, occupied by 3 channels at 4.4 Mbit/s transmission rate, or by 12 channels at 1.1 Mbit/s transmission rate, is sampled at a rate of about 20 MHz and fed into the demultiplexer, which uses a per-channel, "analytic signal" approach. The MCDD contains not only the digital portion of the system but also the analog front end. At the input of the MCDD, an Analog Input Interface is provided that is able to accept the signal at intermediate frequency (140 MHz), to perform the anti-alias filtering and the down-conversion to baseband, so that the final analog-to-digital conversion is done at the Nyquist rate. At the output of the MCDD a Digital-to-Analog Converter is used for the purpose of testing, and allows an oscilloscope to be used to observe scattering diagrams and other significant parameters.

The MCDD tests were performed at INTELSAT, for 6 channels, 3 at 4.4 Mbit/s data rate and 3 at 1.1 Mbit/s data rate, and consisted of: BER versus E_b/N_o, and BER sensitivities to clock frequency offset, carrier frequency offset, baseband signal amplitude variation, and adjacent channel interference (ACI). Figures 10, 11, and 12 show the performance of BER versus E_b/N_o, clock frequency offset, and carrier frequency offset for the 4.4 Mbit/s data rate.

In the Adjacent Channel Interference test, two adjacent channels (upper and lower) are the interfering channels. Measurements are only performed for the center channel. The other two channels are used for producing interference and the MCDD does not demodulate them. The degradation due to the ACI interference is no more than 0.2 dB loss for the 4.4 Mbit/s data rate case.

The BER versus E_b/N_o results are acceptable. However, bit synchronization needs to be improved in the case where E_b/N_o is equal to or less than 8 dB.

234

BaseBand Processor

Under R&D contract to NEC (Japan), INTELSAT developed POC hardware and software for a BaseBand Processor (BBP) for INTELSAT IBS services and for the lower rates in the IDR services [Ref. 8]. The hardware consists of 12 printed wired boards: one TDM/TDMA Converter, one TDMA/TDM Converter, one FDMA Buffer, five for Switch Circuits, and four for the Control Unit; the block diagram is shown in Figure 13. The TDMA/TDM Converter converts the input data format of TDMA to that of TDM. The FDMA Buffers 1 and 2 convert the input data rate to that required in the Switch Module Array (SMA). Switch Circuits perform data rate changes and all switching functions. The TDM/TDMA Converter converts the output data format of TDM to that of TDMA. The Control Unit provides various kinds of clocks, timing signals and read address data for the other subsystems.

The principal functions of the BBP consist of data rate changing; traffic routing at byte level, including Multiplex (TDM-Down), Multi-cast and Distribution, etc.; TDM/TDMA and TDMA/TDM conversions; and Diagnosis.

IBS/IDR BBP tests include communications between the Host Computer and the BBP (Command and Telemetry), data uploading and verification, hardware control (status, switching), switching functions (multi-cast, TDM-down, distribution, data rate changes), and diagnosis function (Column Control Diagnosis and Switching Module Diagnosis). The IBS/IDR BBP performance meets the specifications.

CONCLUSION

OBP systems in several forms are likely to be very significant payload items, and INTELSAT has sponsored their development in the 1980s. The system benefits and constraints are becoming clearer as a result of such work: connectivity, improved link budgets and flexibility are the main objectives; payload constraints include mass, power consumption and reliability with redundancy and diagnostics.

Recent tests at INTELSAT have verified the correct functioning of the On-Board Modem, the FET MSM, the MCDD and the BBP. The Modem and FET MSM are ready for specification development, whereas the MCDD and BBP need to be further developed as engineering models and further tests will be necessary.

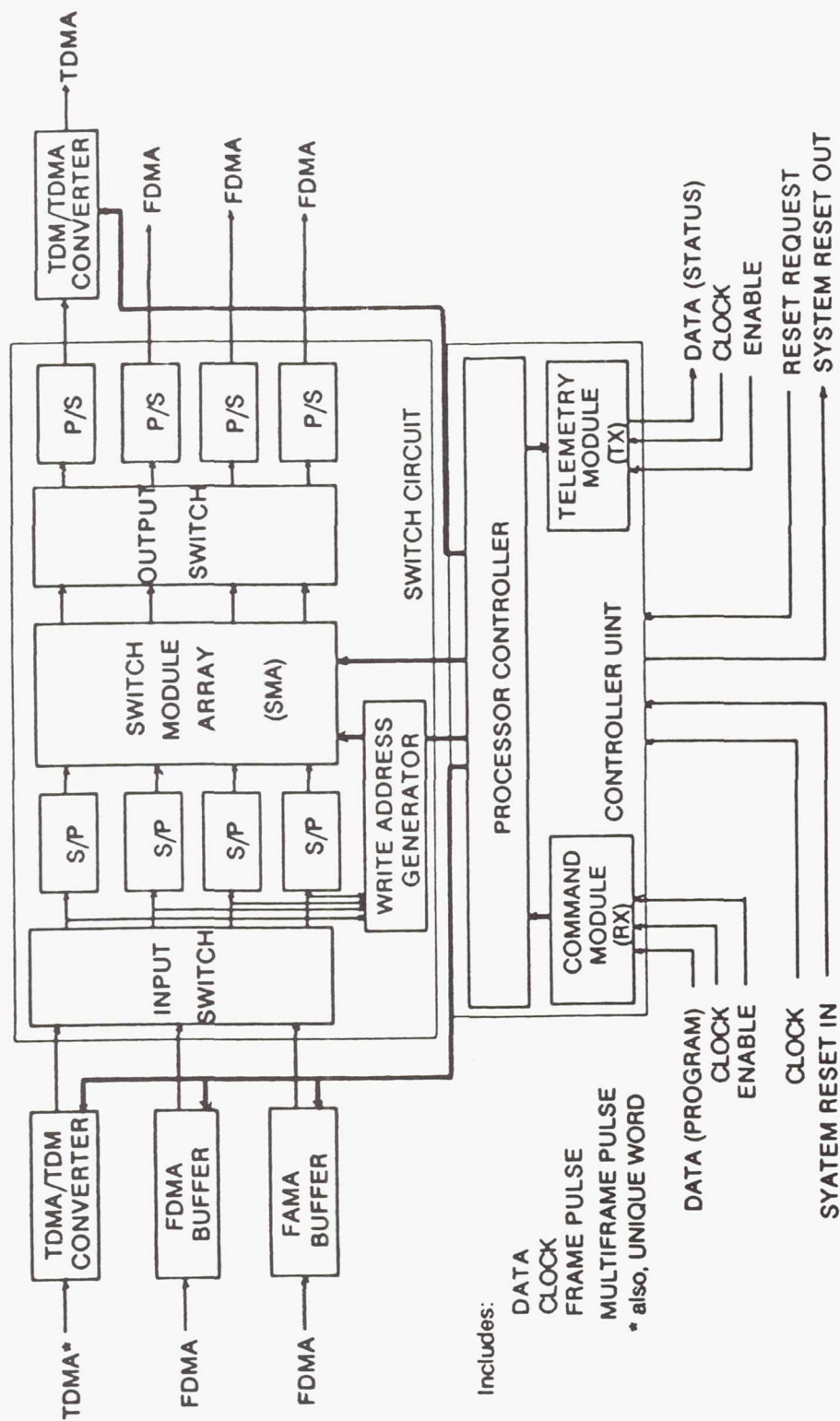A companion poster paper [Ref. 9] shows some more details of the measurements of these OBP subsystems.

Figure 13 IBS / IDR BaseBand Processor (BBP)

236

REFERENCES

1. P. Nuspl, R. Peters, T. Abdel-Nabi and N. Mathews, "On-board Processing for Communications Satellites: Systems and Benefits", INTERNATIONAL JOURNAL OF SATELLITE COMMUNICATIONS, Vol. 5, No. 2, pp. 65-76, Wiley-Interscience, April-June 1987.

2. J. Lunsford, J. Phiel, R. Bedford and S. Tamboli, "SSTDMA System Design", to be published in Special Issue on SS-TDMA, COMSAT TECHNICAL REVIEW, COMSAT Labs, Clarksburg, MD, 1991.

3. Y. Takimoto, S. Kitazume and N. Komiyama, "Satellite Switching Center with Diagnostics for SS-TDMA System", Conference Record, GLOBECOM'86, Houston, Vol. 3, paper 47.1, December 1986.

4. K. Betaharon, S. Kitazume, Y. Takimoto, and N. Komiyama, "Advanced Satellite Switching Center with On-Board Diagnostics", Proc. ICDSC-7, Munich, May 1986.

5. P.V. de Santis, "Non-Regenerative Satellite Switched FDMA (SS/FDMA) Payload Technologies", INTERNATIONAL JOURNAL OF SATELLITE COMMUNICATIONS, Vol. 5, No. 2, pp. 171-182, Wiley-Interscience, April-June 1987.

6. T. Ishizu, Y. Kazekami, H. Sawada and K. Kinuhata, "A Design Approach for On-Board Modems", Conference Record, ICC'86, Toronto, Vol. 3, pp. 1814 - 1820, June 1986.

7. E. Saggese and G. Chiassarini, "SAW and Digital Technologies in Multicarrier Demodulators", INTERNATIONAL JOURNAL OF SATELLITE COMMUNICATIONS, Vol. 7, No. 4, pp. 295-306, Wiley-Interscience, November 1989.

8. T. Kumagai, Y. Takeshima, S. Senba, Y. Ohgushi and P.P. Nuspl, "On-Board Processor for INTELSAT Business Services", Proc. ICDSC-8, Guadeloupe, April 1989.

9. P.P. Nuspl, G. Dong and H.C. Seran, "Laboratory Measurements of On-Board Subsystems", poster paper at Second NASA Space Communications Technology Conference, Onboard Switching and Processing, Cleveland, November 1991.

ADDITIONAL BIBLIOGRAPHY
(chronological)

Y. Kazekami, H. Sawada, K. Kinuhata and P. Nuspl, "Hardware Implementation of On-Board Modem", Conference Record, GLOBECOM'85, New Orleans, December 1985.

P. Nuspl, R. Peters and T. Abdel-Nabi, "On-Board Processing for Communications Satellite Systems: Systems and Benefits", Proc. ICDSC-7, Munich, May 1986.

K. Betaharon, K. Kinuhata, P. Nuspl, and R. Peters, "On-Board Processing for Communications Satellite Systems: Technologies and Implementation", Proc. ICDSC-7, Munich, May 1986.

K. Betaharon, K. Kinuhata, P. Nuspl, and R. Peters, "On-board Processing for Communications Satellite Systems: Technologies and Implementations", INTERNATIONAL JOURNAL OF SATELLITE COMMUNICATIONS, Vol. 5, No. 2, pp. 139-146, Wiley-Interscience, April-June 1987.

F. Ananasso and P. de Santis, "On-Board Technologies for User-Oriented SS-FDMA Satellite Systems", Proc. ICC'87, pp. 8.2.1-8.2.5, Seattle, June 1987.

N. Mathews, "Performance Evaluation of Regenerative Digital Satellite Links with FEC Codecs", SATELLITE INTEGRATED COMMUNICATIONS NETWORKS, North-Holland, 1988.

P. de Santis, "Reconfigurable On-Board Connectivity for INTELSAT Thin Route Services", Proc. ICDSC-8, Guadeloupe, April 1989.

Legend:
ICC   -   International Conference on Communications (IEEE)
ICDSC -   International Conference
              on Digital Satellite Communications

On-Board Congestion Control for Satellite Packet Switching Networks

Pong P. Chu*
Department of Electrical Engineering
Cleveland State University
Cleveland, Ohio 44115

Abstract

It is desirable to incorporate packet switching capability on-board for future communication satellites. Because of the statistical nature of packet communication, incoming traffic fluctuates and may cause congestion. Thus, it is necessary to incorporate congestion control mechanism as part of the on-board processing to smooth and regulate the bursty traffic. Although there are extensive studies on the congestion control for both baseband and broadband terrestrial networks, these schemes are not feasible for space based switching networks because of the unique characteristics of satellite link. In this article, we propose a new congestion control method for on-board satellite packet switching. This scheme takes into consideration of the long propagation delay in satellite link and takes advantage of the satellite's broadcasting capability. It divides the control between the ground terminals and satellite, but distributes the primary responsibility to ground terminals and only requires minimal hardware resource on-board satellite.

1 Introduction

Future satellites are expected to incorporate more on-board processing capability and to support more diversified communication requirements, including integrated data/voice or ISDN (Integrated Service Digital Network) compatible traffic [7, 11]. Since packet switching can obtain more flexibility and efficiency for

---

*The study was performed when the author was visiting NASA Lewis Research Center.

bursty traffic, it is beneficial to incorporate packet-switching capability on-board satellite.

One characteristic of packet switching is that the required bandwidth is allocated on demand rather than on the fixed peak rate [19]. Because of the statistical fluctuation, total incoming traffic may occasionally exceed the capacity of outgoing link even the average incoming volume is within the limit. Thus, it is necessary for a packet switching network to incorporate certain *congestion control* mechanisms to smooth and regulate the bursty traffic. Although this topic is thoroughly studied for terrestrial networks [1, 6, 16], very little is known for the satellite based networks. This paper overviews the congestion control for both baseband and broadband terrestrial networks, proposes a method that is suitable for the unique operational environment of satellites and suggests promising directions for future development.

The remaining paper is organized as follows: Section 2 introduces basic concepts related to the congestion control; Section 3 overviews the congestion control schemes for baseband and broadband terrestrial networks; Section 4 describes the proposed scheme for satellite based packet switching networks; Section 5 outlines the issues for future investigation and last section summarizes the study.

2 Basic Concepts

To achieve flexibility and efficiency, packet switching systems allocate resources on demand; i.e., no bandwidth will be consumed if the connected link is idle. This feature is essential for bursty traffic, in which the packet arrival rate fluctuates significantly and the peak arrival rate is much larger than the average arrival rate. If the allocated bandwidth is fixed, as in the circuit switching, it has to be equal to the *peak rate* to accommodate the worst case. On the other
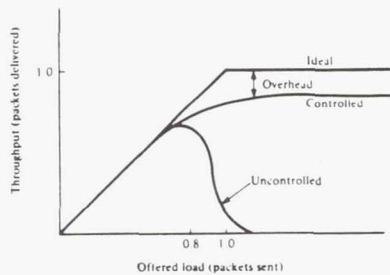
effective throughput vs. offered load
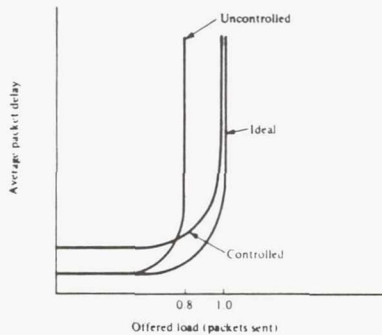
delay vs. offered load

Figure 1: The Effect of Congestion Control

hand, in the packet switching network the allocated bandwidth only needs to be roughly equal to *average rate* of the incoming traffic. Therefore, system resource can be better utilized in packet switching networks.

Due to the statistical fluctuation, packet switching network may suffer potential congestion problem. The volume of total incoming traffic may occasionally exceed the capacity of outgoing link, even the outgoing link can incorporate incoming traffic *statistically* (i.e., in average). Without proper control, the buffer may overflow and certain packets will be lost. These in turn will reduce the effective throughput and introduce long delay. The purpose of congestion control is to provide a mechanism to smooth out fluctuation by regulating the incoming traffic and to prevent severe performance degradation [6]. The typical performance of an ideal system, a controlled system and an uncontrolled system is shown in Figure 1 [19].

2.1 The delay*bandwidth parameter

Although communication networks can be characterized by a wide variety parameters, the term *de-*

*lay*bandwidth* is most essential for congestion control. Delay (d) represents the required time to propagate a packet from source to destination. Bandwidth (B) represents how fast a packet can be transmitted, which normally has a unit of *bits per second*. Delay is fixed for all technology, which is determined by the distance between source and destination and is roughly equal to $\frac{distance}{speed\ of\ light}$. Bandwidth varies with the underlying technology, ranging from 1G *bits/sec* (as in a fiber optic link) to 1K *bits/sec* (as in a phone line).

The bandwidth*delay gives a good indication of the "responsiveness" of the feedback control between source and destination. For example, if destination sends a feedback message at time t_0, the source will receive it at time $t_0 + d$; however, before the source can receive the message, there are already outstanding $d * B$ bits (in the worst case) on their way. For effective congestion control, the scheme needs to consider the current status of the buffer as well as to predict the outstanding traffic already in transmission. In general, designing congestion control schemes is harder for a network with large delay*bandwidth because of the uncertainty associated with the potentially large outstanding traffic.

Delay*bandwidth may vary drastically from one network to another Consider following examples:

- typical LAN (in which $B = 10$ *Mbits/sec*, $d = \frac{1\ km}{c}$): 50 bits.

- a   200   mile   T1   line (in which $B = 1.544$ *Mbits/sec*, $d = \frac{320\ km}{c}$): 2500 bits.

- a 200 mile optical fiber link (in which $B = 1$ *Gbits/sec*, $d = \frac{320\ km}{c}$): 1.6 Mbits.

- a low bit rate satellite beam (in which $B = 64$ *Kbits/sec*, $d = 125$ *msec*): 8000 bits.

- total uplink capacity of a satellite (in which $B = 512$ *Mbits/sec* [11], $d = 125$ *msec*): 64 Mbits.

As we can see, the delay*bandwidth of a satellite based network is significantly larger than its terrestrial counterpart.

2.2 An Unified Permit-Bank Model

The key of congestion control is to develop a mechanism to regulate the rate of incoming traffic so that the outgoing link will not be overwhelmed. In this
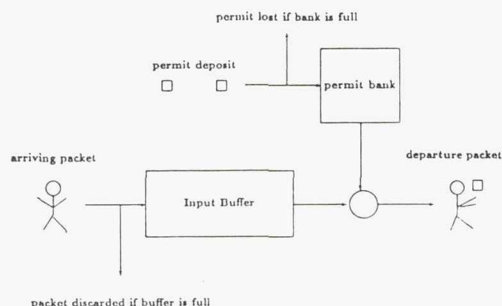
permit lost if bank is full

permit deposit

permit bank

arriving packet

departure packet

Input Buffer

packet discarded if buffer is full

Figure 2: Diagram of Permit Bank

article, we use a simple "permit bank" model (which is also known as *leaky bucket* [4] or *token bank* [2]) to illustrated various congestion control schemes. The basic diagram of this model is shown in Figure 2. There is a buffer and a permit bank for every incoming link. The arriving packet needs to obtain a "permit" before it can be transmitted; otherwise it has to be queued in buffer or discarded. The major design issues is to develop a mechanism to "deposit" permits into the bank and to determine the size of the bank.

3 Current Approaches for Terrestrial Networks

Congestion control for terrestrial networks has been studied extensively [1, 6]. It can be basically divided into two classes: *reactive control* and *preventive control* (or *closed-loop control* and *open-loop control*). Reactive control is normally used for traditional baseband network, in which the delay*bandwidth is relative small. Preventive control is primarily aimed at the fiber-optic based BISDN (Broadband ISDN) ATM (Asynchronous Transfer Mode), in which the high data rate of optical fiber significantly increases the value of delay*bandwidth. The following two subsections outline the schemes used for the two classes. We concentrate more on the preventive control because of the resemblance of ATM networks and satellite networks (the large delay*bandwidth).

3.1 Reactive control

Reactive congestion control is invoked upon the detection of congestion. It normally depends on the feedback mechanism that sends control or status information back to the source. The source then reduces the input rate accordingly.

The most commonly used method is *sliding window*. It can be explained by permit bank model of section 2.2. In this scheme, permit deposit is controlled by the destination. A departing packet will return the permit to destination. Depending on its buffer availability, the destination can either hold the permit or deposit the permit back (by an acknowledge packet) to the source. This holding mechanism implicitly controls the source's packet departure rate.

The size of the permit bank corresponds to the size of the window. It can be pre-determined (such as in internet) or adaptive (such as the *pacing* scheme in SNA) [17, 14]. The adaptive method relies on certain feedback information from network, such as the total round trip delay or number of hops. This information gives an indication of network's status and is carried by acknowledge packets.

The reactive control, in general, is not effective for network with large delay*bandwidth because of the relative slow feedback. By the time the feedback information reaches the source and rate control is triggered, it may be already too late to react effectively. Thus, the proposed high speed BISDN ATM network normally does not employ reactive control schemes.

3.2 Preventive control

Preventive control does not employ feedback information. Instead of reacting to the occurrence of congestion, it tries to prevent the network from reaching an unacceptable level of congestion. Various preventive control schemes are proposed for ATM networks and are an important research issue for BISDN [1].

Preventive control for ATM networks includes two major parts: *admission control* and *bandwidth enforcement*. Admission control determines whether to accept or reject a new connection at the time of call setup, and bandwidth enforcement monitors individual connections to ensure that the actual traffic flow conforms with that reported at call establishment.

admission control Admission control decides whether to accept or reject a new connection based on whether the required performance can be maintained. When a new connection is requested, the network first

241

examines the required service and traffic characteristics as well as the network's current load and status, and then determines whether or not to accept the new connection [8].

To effectively utilize this scheme, three major issues need to be resolved:

- the choice of traffic descriptors.
- the decision criteria.
- the effects of traffic descriptors on the network performance.

Because of the diversity of the expected ISDN traffic, it is very difficult to develop a model to predict the network's performance based on limited information obtained during call setup.

bandwidth enforcement Bandwidth enforcement mechanism is used to monitor the incoming traffic to ensure that the flow conforms with that specified at call establishment. The *leaky bucket* and its variations are the most commonly used methods [4, 2, 15, 18]. The permit bank model also can be used for leaky bucket. In this case, the permit is automatically deposited in a fixed rate, and will be discarded if the permit bank is full. The deposit rate and the size of the bank are determined at the call establishment. The rate corresponds the average rate incoming traffic and the size of the bank indicates the allowed *"burstiness factor"* of the transmission.

The input buffer can be used to queue the incoming traffic exceeding the designated burstiness factor. It provides a better control of the trade-off between packet waiting time and packet loss probability.

4   Congestion Control for On-board Satellite Switching

The packet switching satellite is like a single gigantic concentrator (or a switch) in the sky. Its characteristic is quite unique and thus needs new congestion control schemes. In following subsections, we first examine the differences between satellite networks and terrestrial networks and then describe the proposed scheme.

4.1   Comparisons between terrestrial network and satellite

From congestion control point of view, satellite based packet switching differs from terrestrial network in several aspects: propagation delay, topology complexity and operation environment. Their differences and the implication for congestion control are discussed as follows.

propagation delay Communication satellites normally station at geostationary orbit, approximately 22,300 miles from the surface of the earth. This orbit makes the propagation delay about 125 msec, which is much longer than any terrestrial link. Consequently, as we have seen in section 2.2, satellite link has the largest bandwidth*delay. Because of the sluggish feedback, reactive control similar to sliding window tends to be less effective.

topology complexity Terrestrial networks are normally composed of a number of switching nodes interconnected by various types links. On the other hand, a satellite network contains only *one* switching node, which can be accessed by all users.

The implication of a single node is twofold. First, network congestion status can be easily obtained by observing the available buffer space. Second, satellite can easily broadcast this information to all the ground terminals (this is contrary to terrestrial networks, in which , it is extremely hard to obtain its accurate "global" status and distribute to all the users). From this point of view, it is desirable to maintain certain degree of feedback control in a satellite based network.

operation environment The operation environment for satellite is rather harsh. The devices need to tolerate high temperature and radiation, and their volume, weight and power consumption are severely constrained. Also, maintenance and updating are extremely difficult.

From this point of view, the congestion scheme should distribute more functionality to the ground terminals, and make the space-based segment simple, flexible, and robust.

4.2   Proposed scheme

Because of satellite's unique characteristics, congestion control schemes for terrestrial networks are not feasible for space based packet switching. Sliding window method is not effective because of the long propagation delay. Furthermore, its fairly sophisticated protocol, such as time-out mechanism,

makes on-board implementation impractical (considering that there are about 8000 ground terminals).

Although preventive control is appealing, it is not completely satisfactory because of the potential low utilization. It comes from the fact that the switching node needs to make prediction by *a priori*, which only provides limited information. To guarantee the quality of the service (i.e., low packet loss probability), the admission tends to be conservative and cannot fully utilize the resource. This is less a problem for optical fiber based link since the bandwidth is relatively inexpensive, but is severe for satellite link in which bandwidth is still a precious resource.

The proposed congestion control method is a scheme that combines reactive control and preventive control. This scheme takes into consideration of the long propagation delay in satellite link and takes advantage of the satellite's broadcasting capability. It divides the control between the ground terminals and satellite, but distributes the primary responsibility to ground terminals and only requires minimal hardware resource on-board satellite. It takes advantage of the broadcasting capability of the satellite to overcome the long propagation delay.

This scheme include three major parts: admission control, bandwidth enforcement and a *"group feedback"* mechanism. The first two parts are similar to those in ATM congestion control. However, the two key parameters, deposit rate and bank size, can be updated dynamically according to the buffer availability of dowlinks. To achieve this, the satellite continuously broadcasts its *"congestion status"* and the ground terminals adjust their parameters accordingly. Note that, unlike sliding window method, satellite does not exchange information with ground terminals in an individual basis (so we coin the scheme as group feedback). The detailed description of the proposed scheme is described as follows:

call establishment At the call setup, the admission control will be exercised. The satellite examines the required service, traffic parameters and its own load status and decides whether to accept or reject the call. The initial value of deposit rate and bank size are determined at this time. Although these values may be updated and modified later via group feedback, they establish a "reasonable" basis for future operation and can avoid extreme, unpredicted fluctuation.

Since the deposit rate and bank size can be altered later, the admission control does not need to be very accurate and the task is considerably easier than that of ATM.

group feedback The feedback control is distributed between ground terminals and satellite. The satellite continuously broadcasts the congestion status, and ground terminals react and adjust their parameters accordingly. If the traffic is heavy and congestion is expected, the ground terminals should lower their deposit rate and bank size. In general, reducing the bank size can smooth out burstiness in a short term and reducing deposit rate will regulate and shape the traffic in a long term.

The simplest mechanism to throttle the input traffic is to lower the rate and size of all active terminals in the same proportion. A more sophisticated mechanism should examine various traffic types and their required quality of service, and then respond accordingly (e.g., reduce less for the terminal with real-time traffic). In the ideal situation, the feedback congestion control should be operated on a "continuous" basis instead of just "on or off" by some trigger. In this case, the parameters will be reduced gradually and the performance degradation will be graceful.

4.3 Potential advantages

There are several benefits to implement the proposed congestion control scheme in a space based switching network:

simple implementation The requirements for the space-based segment is relatively simple. Its major tasks are to monitor the incoming traffic and collect the statistics, and to estimate the congestion status from the data collected and broadcast it back to ground terminals.

flexibility Since the execution of group feedback is primarily done at ground segment, the design can be modified or updated without changing satellite's configuration. This feature is desirable since many factors, such as the type of services, its traffic characteristics etc., are not well understood and thus no precise design decision can be made. The proposed scheme allows the exact design to be incorporated into the ground segment when the issues are better understood, and also gives room to add new features in the future.

243

robustness Unlike the sliding window method, the group feedback scheme regulates incoming traffic on a statistical basis. Thus, small fluctuation can be tolerated. Occasional loss of status information will only slightly degrade the performance and will not cause drastic effects.

5 Issues for Future Study

This article only gives a preliminary investigation. To determine the feasibility and effectiveness of the proposed scheme, many issues needs to be carefully studied and many features needs to be fine tuned. The following topics are essential and need to be further studied:

traffic characteristics Because of the statistical nature of the proposed scheme, the understanding of incoming traffic characteristics plays an important role. To accurately describe ISDN traffic will be difficult due to the diversity of the services provided. Models for various types of traffics, such as voice, still video, continuous video etc., have been proposed [1]. However, model for heterogeneous traffic (traffic with mixed typed services) is hardly developed and needs further investigation.

congestion status Recall that the satellite continuously broadcasts congestion status information. This information should take into consideration of:

- current satellite load (the available buffer space the number of active connections etc.).

- the expected incoming traffic in next 125 msec (the propagation delay between ground terminals and satellite).

- the expected incoming traffic after 125 msec.

Note that the second part represents the amount of traffic arriving at satellite before ground terminals can receive the congestion information.

This congestion status can be derived either by conventional *statistic approach*, or by more radical approaches, including *neural network* and *fuzzy logic* [3, 10, 9]. Both neural network and fuzzy logic represent *model-free estimation*, which can incorporate more uncertainty and are more fault tolerant [13].

group feedback control The detailed mechanism for group feedback control for ground segment needs to be investigated. The major tasks are:

- to determine the maximal, optimal and minimal values of deposit rate, bank size and buffer size for various types of service.

- to develop a method to gracefully reduce these values if needed.

call setup Admission control should be done during the call establishment. Since its nature is similar to that of BISDN ATM networks, vast amount of research in ATM networks can be applied [1] and therefore is less crucial.

evaluation The final design should be evaluated by analytical queuing model or by simulation. Its performance and effectiveness should be determined by carefully examining following parameters:

- packet loss probability due to the satellite congestion.

- packet loss probability due to the satellite congestion *plus* ground buffer overflow.

- packet delay and jitter due to satellite link.

- packet delay and jitter due to ground buffer *plus* satellite link.

- utilization of the satellite link.

6 Summary

In this article, we overview the congestion control for terrestrial networks, compare the operation environment between terrestrial networks and satellite networks, and proposes a method that is suitable for satellite based packet switching networks. The proposed scheme employs a permit bank model, and uses a global congestion information broadcasted from satellite to adjust the deposit rate and bank size of ground terminals. This article describes the basic operation of this scheme and suggests the directions and issues for future study.

244

Acknowledgement

References

[1] J. J. Bae and T. Suda "Survey of Traffic Control Schemes and Protocols in ATM Networks" *Proceedings of IEEE* vol 79(2), Feb. 1991, pp. 170–189.

[2] A. W. Berger "Performance Analysis of a Rate-Control Throttle where Tokens and Jobs Queue" *IEEE JSAC*, vol 9 (2), Feb. 1991, pp. 165–170.

[3] E. A. Bobinsky "Traffic and Congestion Modeling of a Proposed Switching Satellite Achitecture", Master Thesis, Cleveland State University, 1991.

[4] M. Butto, E. Cavallero and A. Tonietti "Effectiveness of the Leaky Bucket Policing Mechanism in ATM Networks" *IEEE JSAC*, vol 9 (3), April 1991, pp. 3435–342.

[5] L. Dittmann, S. B. Jacobsen and K. Moth "Flow Enforcement Algorithm for ATM Networks" *IEEE JSAC*, vol 9 (3), April 1991, pp. 343–350.

[6] M. Gerla and L. Kleinrock "Flow Control: A Comparative Survey" *IEEE Trans. on Communications* vol COM-28(4), April 1980, pp. 553–574.

[7] J. L. Harrold, J. M. Budinger, and G. H. Stevens "On-board Switching and Processing" *Proceedings of IEEE* vol 78(7), July 1990, pp. 1206–1213.

[8] Z. Hass and J. H. Winters "Congestion Control by adaptive admission" *IEEE INFOCOM*, 1991, pp. 560–569.

[9] A. Hiramatsu "ATM Communications Network Control by Neural Networks" *IEEE Trans. on Neural Networks*, vol 1(1), Mar. 1990, pp. 122–130.

[10] J. M. Holtzman "Coping with Broadband Traffic Uncertainties: Statistical Uncertainty, Fuzziness, Neural Networks" *IEEE GLOBECOM*, 1990, pp. 7–11.

[11] W. D. Ivancic and M. J. Shalkauser "Destination Directed Packet Switch Architecture for a 30/20 GHz FDMA/TDM Geostationary Communication Satellite Network", in this *Proceeding*.

[12] V. Jacobson "Congestion Avoidance and Control" *ACM SIGCOMM*, 1988, pp. 314–329.

[13] B. Kosko *Neural Networks and Fuzzy Systems*, Prentice Hall, Englewood Cliffs, NJ, 1992.

[14] D. Mitra and J. B. Seery "Dynamic Adaptive Windows for High Speed Data Networks with Multiple Paths and Propagation Delays" *IEEE INFOCOM*, 1991, pp. 39–48.

[15] E. P. Rathgeb "Modeling and Performance Comparison of Policing Mechanisms for ATM Networks" *IEEE JSAC*, vol 9 (3), April 1991, pp. 325–334.

[16] J. Robison et al. "Congestion Control in BBN Packet-Switched Networks" *Computer Communication Review*, Jan. 1990, pp. 76–90.

[17] M. Schawartz *Telecomunication Networks: Protocols, Modeling and Analysis*, Addison-Wesley, Reading, MA, 1987.

[18] M. Sidi et al. "Congestion Control Through Input Rate Regulation" *IEEE GLOBECOM*, 1989 pp. 1764–1768.

[19] W. Stallings *Data and Computer Communications, 3rd ed.*, Macmillan Publishing Company, New York, 1991.

Page intentionally left blank

A B-ISDN-COMPATIBLE MODEM/CODEC*

F. Hemmati and S. Miller
COMSAT Laboratories
Clarksburg, Maryland 20871-9475

ABSTRACT

Coded modulation techniques for development of a B-ISDN-compatible modem/codec are investigated. The selected baseband processor system must support transmission of 155.52 Mbit/s of data over an INTELSAT 72-MHz transponder. Performance objectives and fundamental system parameters, including channel symbol rate, code rate, and the modulation scheme, are determined. From several candidate codes, a concatenated coding system, consisting of a coded octal phase shift keying modulation as the inner code and a high-rate Reed-Solomon as the outer code, is selected, and its bit error rate performance is analyzed by computer simulation. The hardware implementation of the decoder for the selected code is also described.

INTRODUCTION

Current INTELSAT V/V-A time-division multiple access (TDMA) links use quadrature phase shift keying (QPSK) modulation with a transmission rate of 60 Msymbol/s. The INTELSAT V/V-A system has a transponder frequency spacing of 80 MHz and a usable bandwidth of 72 MHz per transponder. With forward error correction (FEC) coding of rate 7/8, the bandwidth efficiency of the QPSK TDMA system is about 1.31 bit/s/Hz of the allocated bandwidth.

A recently developed coded octal PSK (COPSK) modem with a transmission rate of 60 Msymbol/s supports an information rate of 140 Mbit/s. This modem has been field tested to demonstrate restoration of the TAT-8 fiber optic cable by satellite. The implemented 140-Mbit/s modem/codec consists of a 16-state trellis code of rate 7/9 and an OPSK modem (ref. 1). The bandwidth efficiency of the 140-Mbit/s COPSK system is 1.57 bit/s/Hz of the allocated bandwidth, which is an improvement of 33 percent over the QPSK TDMA system.

The synchronous optical network (SONET) is a family of interfaces primarily for use in optical networks. The SONET standard is designed to specify how optical signals would be transported between a number of different vendors' equipment and networks. This standard, along with several other specifications, provides an interface to broadband integrated services digital networks (B-ISDNs). A B-ISDN provides broadband services such as broadcast TV, high-definition TV, and transmission of database files at a high data rate. The standard line bit rate of OC-3 (optical carrier level 3) in the SONET hierarchy is 155.52 Mbit/s, which equals the standard bit rate for B-ISDN.

Given the high reliability of satellites, it is advantageous for network operators to have available economical, B-ISDN-compatible links via the INTELSAT system using only one 72-MHz transponder. Such satellite links can interconnect B-ISDN networks and provide early introduction of this service. Satellites also offer worldwide connectivity and, if B-ISDN is to prosper in many areas of the world, then satellite support is

---

necessary. Satellites can also act as a "safety valve" in optical fiber networks. Note that, in the case of fiber failure or network congestion, traffic can be routed through a satellite channel on a demand-assigned basis.

This paper investigates the design, performance, and implementation of a B-ISDN-compatible modem/codec. System performance objectives and parameters are briefly reviewed, and OPSK is selected as the modulation format because of its constant envelope and the high power efficiency achievable when combined with an appropriate code. Coded modulation techniques, including Ungerboeck codes and Imai-Hirakawa codes, are then briefly reviewed. The selected code is an Imai-Hirakawa code, with block/convolutional component codes concatenated with a high-rate Reed-Solomon (RS) code in order to achieve high integrity. The hardware implementation and bit error rate (BER) performance of the proposed code over an additive white Gaussian noise (AWGN) channel and a typical INTELSAT V nonlinear channel are presented, and the flexibility of the decoder for transmission of 140 Mbit/s is examined in detail.

SYSTEM PARAMETERS AND PERFORMANCE OBJECTIVES

For a given information bit rate, fundamental parameters for a coded modulation system include the channel symbol rate, the code rate (R), the modulation scheme, and the expected error performance. A transmission symbol rate of 60 Msymbol/s is preferred because the QSPK TDMA system is operating at this rate, and hence the available subsystems, such as transmit and receive filters and equalizers, can also be used in the B-ISDN system, thus reducing the overall unit cost. Also, high-level modulation schemes are sensitive to phase noise. The main sources of phase noise are the group delay distortion of pulse-shaping and satellite-multiplexing filters, AM/AM and AM/PM nonlinearities, residual phase modulation in high-power amplifiers (HPAs), and carrier phase noise. Group delay distortion of filters can be almost perfectly equalized when the channel symbol rate is 60 Msymbol/s or less; however, an equalizer for higher channel symbol rates might not perform as well.

International standards for the error performance of B-ISDNs are not yet available. However, it is expected that a BER of approximately 10^{-10} for 90 percent of the available time will be adapted by the CCITT for broadband service.

Extensive link analysis, laboratory hardware measurements, and field trials have indicated that the INTELSAT V transponders are primarily interference limited, not thermal noise limited, when supporting very high data rate services (ref. 2). Bandwidth efficient coded modulation schemes, suitable for high-speed implementation, do not afford sufficient power efficiency for correcting all the errors caused by the link interferences. The BER vs E_b/N_o performance curves exhibit unresolvable errors at bit error rates of 10^{-7} and less. Therefore, a concatenated coding system, consisting of a power- and bandwidth-efficient inner code and a high-rate RS outer code was considered for achieving the required BER performance of 10^{-10}. Performance objectives and system parameters for the B-ISDN codec/modem are summarized in Table 1.

CANDIDATE MODULATION SCHEMES

The transmission of 155.52 Mbit/s over INTELSAT V transponders requires a bandwidth efficiency of about 1.94 bit/s/Hz, or an improvement of 48 percent over the QPSK TDMA system. Because the required bandwidth efficiency cannot be achieved by QPSK modulation, higher level modulation schemes such as OPSK or 16-ary signal constellations must be considered.

Candidate modulation schemes for the B-ISDN channel include OPSK, 16-ary PSK, and 16-ary quadrature amplitude modulation (QAM).

OPSK modulation, together with a suitable code, can achieve good power and bandwidth efficiency over the satellite channels. Sixteen-ary PSK is also a bandwidth-efficient modulation scheme; however, it has not yet been implemented for high-speed applications. The performance of 16-ary PSK modulation is very sensitive to phase noise, and its demodulator requires fine resolution for distinguishing between the 16 points closely packed on the circumference of a circle. Moreover, the complexity of the synchronization circuits for symbol timing and carrier and clock recovery is greater than that for the OPSK demodulator.

248

Table 1. B-ISDN-Compatible Codec/Modem Performance Objectives and System Parameters

Nominal Information Rate	155.52 Mbit/s
Nominal Channel Symbol Rate	60 Msymbol/s
Nominal Code Rate	$\approx 5/6$
Mode of Operation	Continuous
Satellite Transponder	INTELSAT V, V-A, or VI
Usable Transponder Bandwidth	72 MHz
Allocated Bandwidth	80 MHz
Coverages	Hemispheric, Zonal
BER Goals:	
IF Loopback	Better than 10^{-10} at $E_b/N_0 = 11$ dB
Typical Nonlinear Channel	Better than 10^{-10} at $E_b/N_0 = 14$ dB
Energy Dispersal	Scrambling, in accordance with CCIR Rec. 359-3

Sixteen-ary QAM is a power- and bandwidth-efficient modulation scheme, but is most suitable for linear channels. For the present application, the earth station HPAs, and particularly the satellite traveling wave tube amplifiers (TWTAs), must operate in the nonlinear region near their saturation point. Therefore, the BER performance of the 16-ary QAM is not expected to meet the system specifications.

Based on the above factors, OPSK appears to be the only viable candidate modulation method. Therefore, coded OPSK modulation techniques are examined in the next section.

COPSK MODULATION TECHNIQUES AND THE SELECTED CODES

The area of power- and bandwidth-efficient coded modulation techniques has been of great research interest for several years. In addition to the class of coded continuous-phase frequency shift keying (CPFSK) modulation schemes, research in this area has focused in two closely related areas now known as Ungerboeck codes and Imai-Hirakawa codes. CPFSK modulation schemes, which include the class of multi-h codes, are not suitable for the present application because of their low power efficiency and the unmanageable complexity of the modem and codec hardware implementation operating at the required speed.

UNGERBOECK CODES

In 1982, Ungerboeck (ref. 3) introduced the concept of "set partitioning" and applied this idea to constructing bandwidth-efficient trellis codes. A properly designed trellis code can provide significant coding gain over an uncoded modulation system. Coding gains of about 4 to 5 dB can readily be achieved with a low- to moderate-complexity decoder, at the expense of a more complicated modem.

The hardware implementation complexity of the decoder for Ungerboeck codes depends on the number of encoder states and the code rate, $R = k/n$. The code rate must be selected to minimize the complexity of branch metric computations.

IMAI-HIRAKAWA CODES

The multilevel coding method proposed in 1977 by Imai and Hirakawa (ref. 4) is convenient for high-speed implementation and, for a selected modulation signal space, allows a wide range for the code rate.

In a multilevel/phase signal space, the Euclidian distances between a particular signal point and the remaining points in the signal set are not equal. Since the distance between adjacent signal points is much smaller than the maximum distance between elements in the signal space, more code redundancy must be allocated for encoding the adjacent points. Similarly, the information bits that distinguish between signal points which are far from each other can be encoded by a high-rate code or remain uncoded.

The structure of the class of codes known as the Imai-Hirakawa codes is based on the above concept, and they are generated by several encoders of various rates, as indicated in Figure 1. Low-rate codes are used for encoding the adjacent symbols, and high-rate codes are selected for encoding signal points located a large
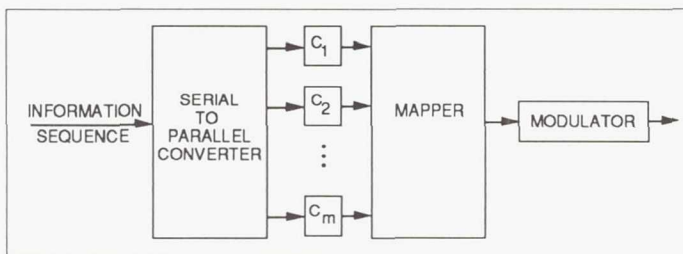
Figure 1. A Coded Modulation System Using
Imai-Hirakawa Codes

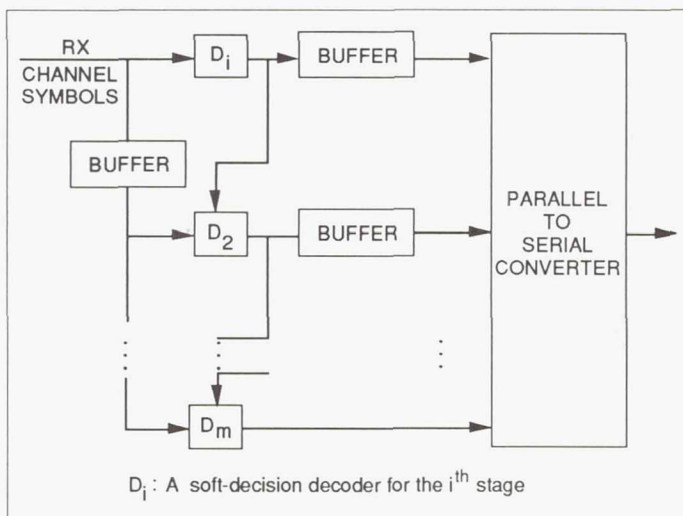

D_i : A soft-decision decoder for the ith stage

Figure 2. Multistage Decoding Procedure for
Imai-Hirakawa Codes

distance from each other. For a signal space with 2^m elements, $C_i = (n, k_i, d_i)$, $1 \leq i \leq m$, constitutes the set of m component block codes with a common code length n, where $R_i = k_i/n$ defines the code rate and d_i denotes the minimum Hamming distance for the ith code. Both block codes and convolutional codes can be used as the component codes.

Generalized versions of the Imai-Hirakawa codes have been designed by Ginzburg (ref. 5) and Sayegh (ref. 6). In particular, Sayegh described the efficient multistage decoding procedure of Figure 2, in which the most error-prone information bits are estimated first, using *a posteriori* probabilities based on the received channel symbols and the code structure. These estimates are then used in later decoding stages to estimate the successively less error-prone information bits.

Power- and bandwidth-efficient Imai-Hirakawa codes can be readily constructed by using the available optimum block/convolutional codes as the component codes. More importantly, available high-speed decoders can be used in the multistage decoding procedure.

Among several candidate coded OPSK modulation schemes, a block/convolutional Imai-Hirakawa code of rate 13/15, affording an asymptotic coding gain of 3.58 dB over uncoded QPSK, was selected as the inner code (ref. 7). While other coded QPSK modulation schemes potentially might yield a higher power/bandwidth efficiency, this particular code was chosen mainly because of its simplicity of hardware implementation.

The encoder for the block/convolutional code of rate 13/15 consists of three component encoders whose outputs are arranged in a 3-row by 15-column array. The first row of the array is a block of 15 consecutive encoded bits, generated at the output of a punctured convolutional encoder of rate 2/3 and constraint length 7. The second component code is a single-parity check code of length 15 and rate 14/15, and the third row of the array is a block of 15 uncoded information bits, which is a codeword in the universal (15, 15, 1) code of rate 1. The 3 bits in each column of the array specify one of the points in the OPSK signal space.

After analyzing the burst error statistics of the inner code, an 8-symbol error correcting (255,239) RS outer code was found suitable for achieving the required BER performance. The overall rate of this concatenated coding system is approximately 13/16 and requires a channel symbol rate of 64 MHz for supporting 155.22-Mbit/s data and synchronization overhead bits.

The required channel symbol rate can be reduced to 62.4 MHz by using a (15, 15, 1) universal code instead of the (15, 14, 2) parity check code in the second encoding stage. In this case, the inner code operates at 58.5 MHz. Finally, a 140-Mbit/s codec can be realized if the inner code of rate 13/15 is concatenated with a (195, 175) 10-error-correcting shortened Reed-Solomon code requiring a channel symbol rate of 60 MHz.

PERFORMANCE ANALYSIS

The BER performance of the considered concatenated coding system was evaluated by first examining the performance of the inner code, without the RS outer code, under various link conditions. Then, the performance of the concatenated coding system was evaluated under worst-case link conditions.

The BER performance of the OPSK coded modulation of rate 13/15 for transmission of 155.52 Mbit/s was evaluated by computer simulation. The received channel symbols were first quantized by a 64-level quantizer, and then compressed to 3 bits by a nonlinear mapping. Over an AWGN channel and at a BER of 10^{-5}, the BER performance of this quantization scheme is within 0.2 dB of the performance of the coded system using unquantized channel symbols.

BER performance results over an AWGN channel and over a typical INTELSAT V nonlinear channel are shown in Figure 3. The system environment and performance parameters considered in the computer simulations are summarized in Table 2. For the AWGN channel, a coding gain of 1.2 dB over uncoded QPSK is observed at a BER of 10^{-5}. An effective coding gain of about 2.5 dB is expected at a BER of 10^{-8}, which can be obtained by extrapolating the BER performance curve of Figure 3. The 1.08-dB discrepancy between the asymptotic coding gain for this code (3.58 dB) and the effective coding gain of 2.5 dB is due to the adversary path multiplicity in the k = 7 punctured convolutional code of rate 2/3, suboptimum multistage decoding (instead of maximum-likelihood decoding), and 3-bit soft-decision quantization (instead of an infinite number of quantization levels).

At a BER of 10^{-4}, the performance of the single nonlinear satellite channel degrades by about 1.5 dB relative to the performance of the AWGN channel, due to link nonlinearities and intersymbol interference. The BER performance degrades by an additional 1 dB with one entry of co-channel interference (CCI) at a power level of -18.5 dB with respect to the desired channel. The two 60-Msymbol/s adjacent channels located at +80 and -80 MHz relative to the center frequency of the desired channel degrade the BER performance by an additional 0.3 dB.

The BER performance results shown in Figure 3 are obtained by assuming that the group delay distortion of the modem filters and satellite multiplexing filters is ideally equalized. In a real channel, phase noise caused by the carrier oscillator, link nonlinearities, and group delay distortion of filters degrades system performance. At a BER of 10^{-4}, degradations due to unequalized group delay distortion can be as much as 0.8 dB.

The BER performance of the concatenated coding system was also evaluated by computer simulation. The results obtained are shown in Figure 4 for the nonlinear channel with two ACI and one CCI at -18.5 dB. The dashed curves show the BER performance of the inner code, without the RS outer coding. The performance of the concatenated coding system with channel symbol rates of 64 Msymbol/s and

Figure 3. Ber Performance of the Rate 13/15 Code Over the Nonlinear Satellite Channel

Table 2. Summary of System Variables and Assumptions Used in the Computer Simulations

Number of Samples per Symbol	16
HPA Input Backoff	10 dB
Satellite TWTA Input Backoff	2 dB
CCI Level	-18.5 dB
Separation Between Adjacent Channels	80 MHz
Rolloff Factor for Square-Root Nyquist Modem Filters	40%
Ideally Group Delay Equalized Filters	
Perfect Symbol Phase and Symbol Timing	
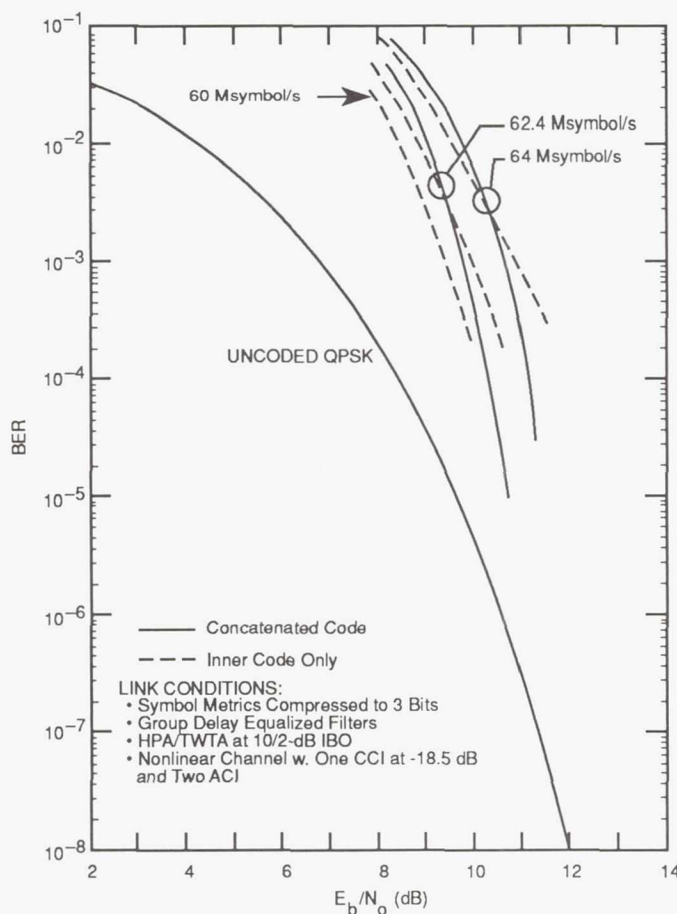
Figure 4. BER Performance of the Concatenated Coding System Over the INTELSAT V Transponders

62.4 Msymbol/s, corresponding to the inner codes with and without parity check coding in the second stage, are depicted as solid curves. Because of its lower channel symbol rate, the 62.4-Msymbol/s system outperforms the 64-Msymbol/s system by about 1 dB. The 62.4-Msymbol/s system is expected to achieve a BER of 10^{-10} at less than 12 dB, which allows sufficient margin for modem implementation loss and the transmission link.

SYNCHRONIZATION

To achieve synchronization and OPSK phase ambiguity resolution, the system employs an innovative digital technique which synchronizes and controls the decoder while simultaneously providing an on-line BER measurement at the codec output. The synchronization method is unique in that the overhead data are periodically inserted into the encoder during transmit data processing. When transmitting data at 155.52 Mbit/s, the overhead data are added at a rate of 1 percent in the form of a single 40-bit data packet inserted into the multistage encoder once after every 100 information blocks have been processed. The encoded overhead packet is largely recognizable at the input to the decoder because the unique word information is transmitted through the uncoded stage of the multistage encoder while the other two stages encode zeros during the overhead packet. This results in a stream of consecutive OPSK symbols that can take on one of two BPSK points in the OPSK signal space. The unique word detection circuit is able to detect the unique word regardless of the phase state the modem is locked in, and, combined with the hard decision data output from the modem, the synchronization circuit effects a digital phase rotation of the data which removes any phase rotation before decoding takes place.

The innovative synchronization scheme offers several advantages over conventional synchronization approaches including elimination of the need for large asynchronous buffers for overhead insertion/removal, insertion of the overhead packet directly into the encoder (overhead packet is directly linked to the structure of the code), removal of the need for a preamble sequence to resolve modem phase ambiguity, elimination of another step in phase-locked loop circuitry, and the side benefit of providing a real-time BER monitoring capability since the overhead data are passed through the decoder before being removed from the data stream.

A number of alarms and operational features from relative CCITT recommendations were incorporated into the proof-of-concept hardware model, which renders the codec design suitable for manufacturing and field deployment. Included among the operational enhancements is the ability of the system to detect an incoming alarm indication signal (AIS) from upstream equipment, the ability to generate an AIS signal when a fatal error in the coded modulation system is present, on-line BER monitoring capability, high BER alarm, standard coded mark inverse (CMI) high speed interface compatibility and plesiochronous/Doppler buffering for operation with satellites in inclined orbits up to 3°. The most prominent maintenance features include the development of universal test equipment that is compatible with all boards and subsystems within the system, baseband loopback capability, clock and data activity detectors and indicators, power supply indicators, and ample test points.

HARDWARE IMPLEMENTATION

Detailed design and construction of the multistage inner coding system has been completed, and hardware design, construction, and test of the RS outer codec is under way and will be completed in 1991.

To implement the concatenated multistage inner code and the RS outer code, the concatenated coding system uses five circuit boards housed in a common chassis with backplane intercommunications. The circuit boards employ an efficient mixture of high-speed emitter-coupled logic (ECL), intermediate speed complementary metal-oxide semiconductor (CMOS) and transistor-transistor logic (TTL) circuitry, and analog phase-locked loop components. The first circuit board implements the high-speed data interfaces to the system, the phase-locked loop timing circuits, and the serial-to-parallel data conversion function. The CMOS/TTL design of the second board executes the outer code RS encoding, decoding, interleaving, and block code synchronization functions. Parallel use of off-the-shelf RS codec chips supports the system's high data rate throughput. A combination of the parallel RS coding devices and auxiliary codeword memory serves to implement the depth of four interleaving that is necessary for the RS code to effectively combat the system's burst errors. An efficient combination of programmable gate array devices and discrete logic and memory comprise the rest of the second board design. The third board uses programmable and discrete CMOS logic to implement the transmit functions for the inner coding system, including multistage encoding, data scrambling, unique word insertion, data frame construction, and interface to the OPSK modulator. The fourth and fifth boards are responsible for the inner code receive-side processing. In particular, the fourth board operates at a symbol rate of 58.5 Msymbol/s and uses high-speed ECL circuitry to perform OPSK phase ambiguity resolution, inner decoder synchronization, Viterbi metric calculation, parity decoding, universal decoding, and overall decoder timing and control. The fifth board uses a parallel array of commercially available Viterbi codec chips with a mixture of CMOS discrete and programmable logic to implement the Viterbi decoding portion of the multistage decoder, the data alignment function, and the Doppler buffering.

The OPSK modulator accepts 3-bit symbols from the encoder and creates one of the eight phase states in the OPSK signal space. The OPSK demodulator receives the incoming IF signal, makes gain adjustments to the signal path, removes the modulation to recover the carrier, generates symbol timing, and demodulates the data into soft decision quadrature baseband streams. In the past, COMSAT Laboratories developed a 180-Mbit/s OPSK modem for use in its 140-Mbit/s rate-7/9 COPSK system (ref. 2). The basic structure for this design was incorporated into the modem design for this project, with enhancements for additional alarms and operational features.

REFERENCES

(1) F. Hemmati and R. J. F. Fang, "Low-Complexity Coding Methods for High-Data-Rate Channels," *Comsat Technical Review*, Vol. 16, No. 2, Fall 1986, pp. 425–447.

(2) H. Layer, J. M. Kappes, and C. B. Cotner, "140-Mbit/s COPSK Modem Laboratory Tests and Transatlantic Field Trials," *COMSAT Technical Review*, Vol. 20, No. 1, Spring 1990, pp. 63–96.

(3) G. Ungerboeck, "Channel Coding With Multilevel/Phase Signals," *IEEE Transactions on Information Theory*, Vol. IT-28, No. 1, January 1982, pp. 55–67.

(4) H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error- Correcting Codes," *IEEE Transactions on Information Theory*, Vol. IT-23, No. 3, May 1977, pp. 371–377.

(5) V. V. Ginzburg, "Multidimensional Signals for a Continuous Channel," *Problemy Peredachi Informatsii (Problems of Information Transmission)*, Vol. 20, No. 1, 1984, pp. 28–46.

(6) S. I. Sayegh, "A Class of Optimum Block Codes in Signal Space," *IEEE Transactions on Communications*, Vol. COM-34, No. 10, October 1986, pp. 1043–1045.

(7) F. Hemmati, "A B-ISDN-Compatible Modem/Codec," IEEE Global Telecommunications Conference, San Diego, California, December 1990, *Conf. Rec.*, pp. 907.4.1–907.4.5.

Page intentionally left blank

FLEXIBLE HIGH SPEED CODEC (FHSC)*

G.P. Segallis and J.V. Wernlund
Harris, Government Systems Sector
Melbourne, Florida 32901

ABSTRACT

This paper describes the on going NASA/Harris FHSC CODEC program. The program objectives are to design and build an encoder decoder that allows operation in either burst or continuous modes at data rates of up to 300 megabits per second. The decoder handles both hard and soft decision decoding and can switch between modes on a burst by burst basis. Bandspreading is low since the code rate is greater than or equal to 7/8. The encoder and a hard decision decoder fit on a single application specific integrated circuit (ASIC) chip. A soft decision applique is implemented using 300K ECL logic which can be easily translated to an ECL gate array.

INTRODUCTION

Principal use envisioned for the technique is to achieve a significant amount of coding gain on high data rate, bandwith constrained channels. Satellite channels and line of sight microwave links up to and including T-4 data rates could benefit from this CODEC.

The Hardware being developed for this program consists of 10 BCH ASIC's, two Flexible High Speed CODEC chassis and a Test and Demonstration chassis. The Flexible High Speed CODEC is a high speed stand alone block encoder/decoder. The decoder provides significant gain with hard decisions alone (up to 4dB) and can utilize soft decision information when available from the demodulator to increase the coding gain by as much as 1.5 dB. The Test and Demonstration chassis provides the link simulator, all control signals and the interface between a serial data generator and the Flexible High Speed CODEC. These chassis along with a commercial bit error rate test set, a PC and a synthesizer are the hardware of this program.

Some interesting aspects of this coding technique include the ability to handle burst (e.g. packet) lengths from 224 bits up, in steps of 32 bits. It may be switched in or out on a burst by burst basis without affecting throughput delay. The interface design allows mating with many forms of M-ary modulation including M=2,4,8 and 16. This paper discusses the FHSC program in general, the approach taken to testing and the hardware. A brief discussion of the program status is presented at the end of the paper.

The HARRIS-NASA BCH CODEC utilizes presolved equations to implement a hard decision, triple error correcting Bose-Chaudhuri-Hocquenghem block codec (ref. 1). This CMOS ASIC contains 18,000 equivalent gates, is packaged in a 132 pin PGA, and consumes 1.5W at +5V. This CODEC will provide up to 4 dB coding gain (see fig. 1) at data rates up to 300 Mbps with low bandspreading. The CODEC may be used in either a full or partial coding scheme and may be interfaced to various types of modems such as m-ary PSK and QAM. The CODEC will correct all patterns of 3 or less errors within a block. In addition, many higher weight error patterns are detected and status lines are provided to the user. Data format may be either continuous or variable length bursts.

Hard Decision Performance of a (256, 224) Code
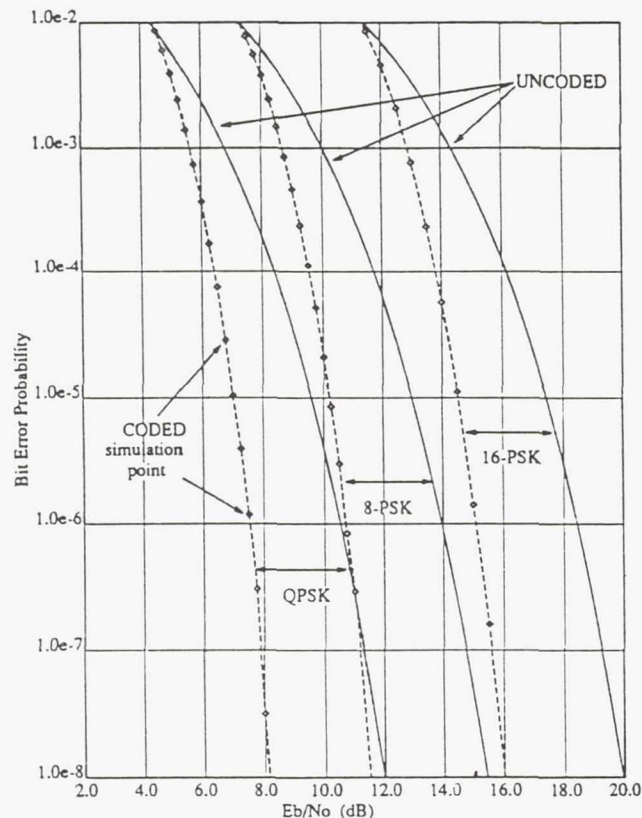
Hard Decision Coding Gain (in dB)						
BER	(256,224) Code			(512,480) Code		
	QPSK	8-PSK	16-PSK	QPSK	8-PSK	16-PSK
10^{-4}	2.0	2.2	2.4	1.8	2.0	2.2
10^{-6}	3.0	3.2	3.4	2.8	3.0	3.2
10^{-8}	3.8	3.9	4.0	3.6	3.7	3.8

Fig. 1 Hard Decision Coding Gains

The CODEC interface is configurable to be either 1, 2, 4 or 8 bits wide. This allows a single symbol wide interface for several modulation types, or symbols may be stacked up to

increase throughput. The interface will operate up to 43 MHz providing a 38 Mbps, 75 Mbps, 150 Mbps, or 300 Mbps data rate for the specified interface widths. The BCH encoder appends 32 parity bits to a data block. Legal block lengths are from 224 to 480 bits in 16 bit increments. Four of the parity bits are user programmable at the encoder and are provided to the user by the decoder on the receive side. This feature can be used to implement a voice or data order wire. The resulting "codeword" is 256 to 512 bits yielding a very high code-rate of 7/8 to 15/16. Note, the 300 Mbps data rate is data only, the CODEC will operate up to 343 Mbps to account for the coding overhead. Bursts longer than 480 data bits are formed by concatenating 2 or more blocks of length 224 to 480 bits. This allows any burst length from 224 on up in increments of 16 bits (see fig. 2).
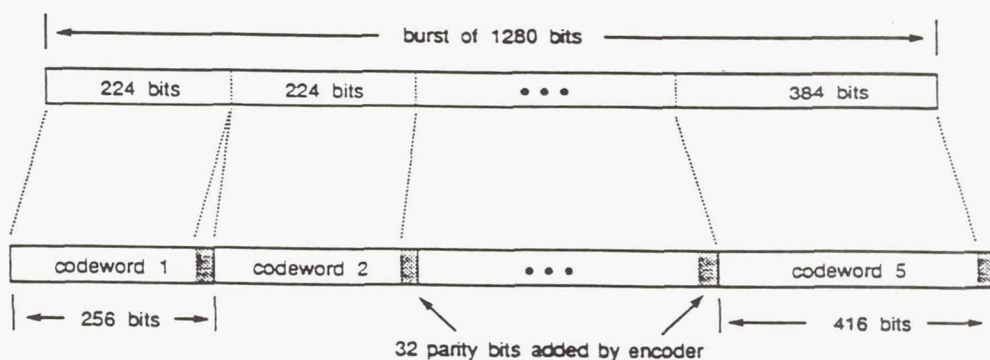
Fig. 2 Concatenated blocks

Block boundaries are either determined by the user and provided to the CODEC or may be internally generated by the CODEC to provide a continuous mode of operation. In this mode the decoder performs a search for codeword position alignment. The decoder can also provide internal demod-carrier phase ambiguity resolution when operating with BPSK, QPSK or 16-ary PSK modems. The codeword position search, phase resolution, as well as lock status are controlled by internal lock circuitry. This lock circuit is externally programmable to allow adaptation to specific performance and acquisition time requirements. All of the lock circuitry inputs and outputs are brought off chip to allow external lock/search control.

Coding may be turned on and off (ie. decoder corrects no errors) on a burst by burst basis without altering the throughput delay. In the uncoded mode data is passed continuously. In this mode block mark is used as a data valid signal and does not transition low until the end of a burst.

The decoder also contains internal fault circuitry providing fault status to the user. Scan chain and mux isolation circuitry allows extensive off-line test of the ASIC.

In addition to correcting any errors found, the decoder also provides their locations to the user. This information is used by the FHSC CHASE soft decision algorithm to further increase coding gain.

FHSC CODEC

The FHSC chassis utilizes 4 of the BCH ASIC's, the Chase circuits and I.O. formatting circuits to implement the flexible high speed CODEC algorithms. It is designed to operate

from DC to 300 megabits while providing as much as 5.4 dB of coding gain (see fig 3). The interface width can be 3 or 4 bits. It can be used in a burst or a continuous mode. And can operate as a hard decision or soft decision CODEC. The CODEC can be switched between coded and uncoded modes on a burst by burst basis. The four user programmable parity bits are brought out from the ASIC's to enable order wire applications. The delay through the encoder/decoder is essentially independent of its modes of operation. There is a slight difference between the delay in 8-ary and any other mode. This is due to the difference between the 8-ary symbol clock and the 3/4 symbol clock used internally.

Soft Decision Performance of a (256, 224) Code
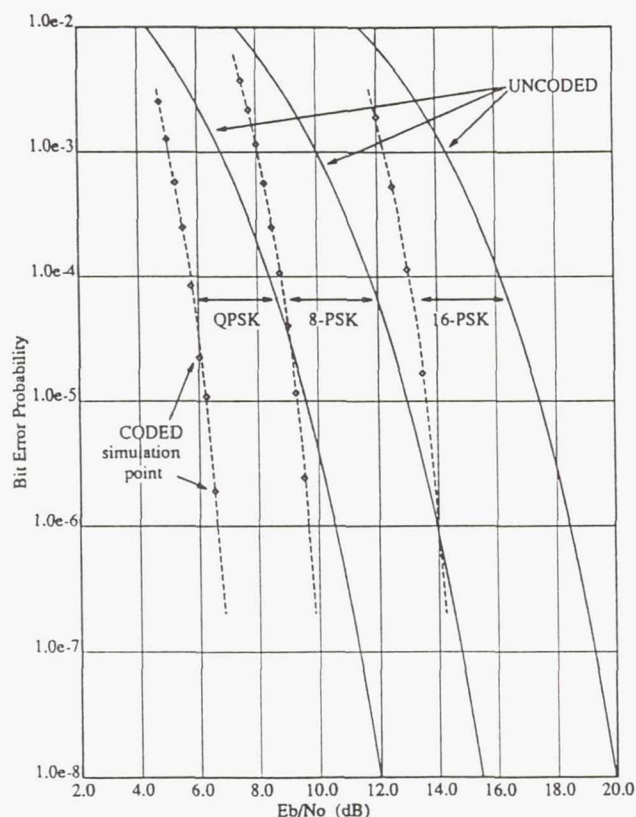
Soft Decision Coding Gain (in dB)						
BER	(256,224) Code			(512,480) Code		
	QPSK	8-PSK	16-PSK	QPSK	8-PSK	16-PSK
10^{-4}	2.7	2.9	3.1	2.5	2.7	2.9
10^{-6}	4.0	4.3	4.5	3.8	4.1	4.3
10^{-8}	4.9	5.2	5.4	4.7	5.0	5.2

Fig. 3 Soft Decision Coding Gains

A single block mark is used to control the all of the circuits with in the FHSC . This signal starts the encoding process the decoding process and controls all of the gated clocks used within the FHSC. In the burst mode this signal is supplied by the user. Because the signal eventually controls the BCH CODEC chips it has the same constraints as it would for the ASIC alone.

In the continuous mode the FHSC generates all block mark signals internally. In this mode the block lengths are forced to 288 bits. An internally generated gated clock is output to the user for clocking data into the encoder. Like the BCH ASIC the FHSC decoder can acquire code word boundaries in all continuous modes. It can also resolve carrier phase ambiguities in hard decision 2-ary, 4-ary and 16-ary modes. Signals out of the FHSC indicate when the decoder is unable to lock or when a fault is detected. In the burst mode a bump phase signal is provided. This signal indicates the BCH decoder would like the TDMA controller to slip the block mark signal one symbol time. The codec is built using 300K ECL, 4 ASIC's and FCT logic. It is designed onto 4 MuPac multi technology wire wrap cards and consumes 350 watts of power.

THE CHASE

The Chase circuits perform the soft decision decoding of the CODEC (ref. 2). The Chase preprocessor circuits identify the three bit positions, with-in a code word, with the poorest statics. These bit positions are then toggled to produce four code words for decoding by the BCH decoder ASIC's (see fig 4). The code word parity is used to determine which set of four code words is generated. A likelihood for each code word is also calculated. These likelihoods are a measure of the correlation between the original received code word and the four generated code words. These likelihoods and the altered bit locations are then stored in FIFO's for use by the Chase post processor circuits. The post processor circuits read the bit locations changed by the BCH decoder chips and calculate a final likelihood. These final likelihoods are then used to select which decoder's output will be selected as the FHSC decoded output. These circuits are all designed to interface with the BCH chip at bit rates up to 342 Megabits.
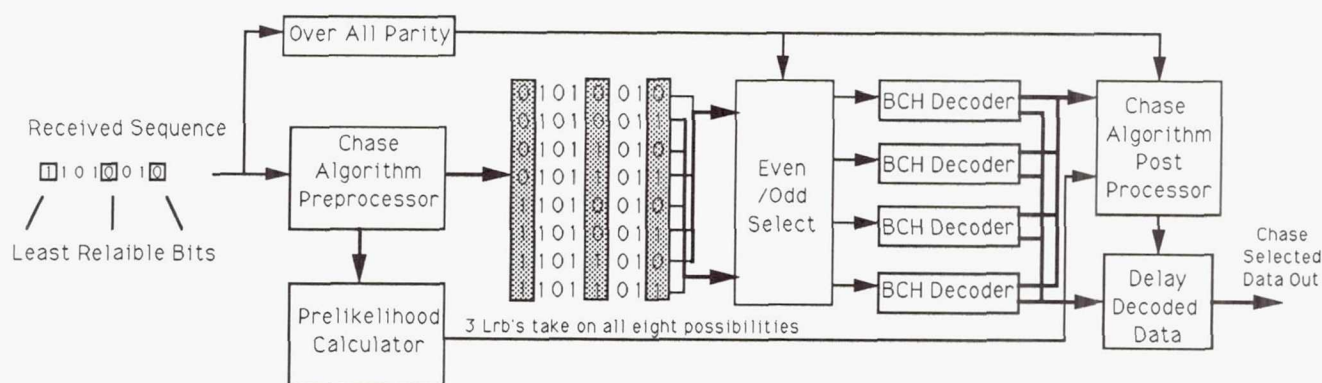
Fig. 4 Chase Algorithm

INPUT OUTPUT FORMATTERS

The FHSC is designed to interface with two symbol widths. They are 3 bits and 4 bits. With-in a four bit interface there may be more than one symbol. This enables the CODEC to be used with BPSK, QPSK, 8-ary or 16-ary symbols. The input formatters first convert the

incoming symbols into four bit nibbles and then into 8 bit words. The four bit nibbles are used by the Chase preprocessor circuits while the 8 bit words are used by the BCH CODEC chip. Input FIFO's are used to buffer the incoming symbols from the nibbles and words used by the preprocessor and the BCH CODEC's. All symbols be they 2-ary, 4-ary, 8-ary or 16-ary are passed through these FIFO's. This avoids phase problems due to delays between the user and the high speed circuits with in the FHSC CODEC.

TDE

The Test and Demonstration equipment is under PC control. It uses test profiles loaded from the PC to generate all the control signals needed by the FHSC. These profiles include number of bursts, burst lengths, modulation modes, coded or uncoded and signal to noise ratios. The burst lengths and number of bursts are used to generate the control signal block mark. The signal to noise ratios and the modulation mode is used to set up the link simulator circuits.

The TDE also serves as a buffer between a serial bit error rate test set and the FHSC chassis. The TDE uses gated clocks to clock serial data to and from the BERT. This data is then formatted into the appropriate symbol width. All clocks needed by the FHSC and BERT's are generated from the synthesizer reference by the TDE.

The link simulator is a full rate simulator. It can generate hard and soft decision symbols at 300 megabits in the 16-ary mode, 225 megabits 8-ary and 150 megabits 4-ary. It is capable of generating noisy symbols for bit error rates from 10E-1 to 10E-10 (see fig 5). Noise profiles are generated by inputting the desired Eb/No and the modulation mode to a noise generator program resident in the test software. Current modulation profiles include 4-ary, 8-ary and 16-ary PSK signals. Many other modulation modes could be tested by modifying the noise generator software.
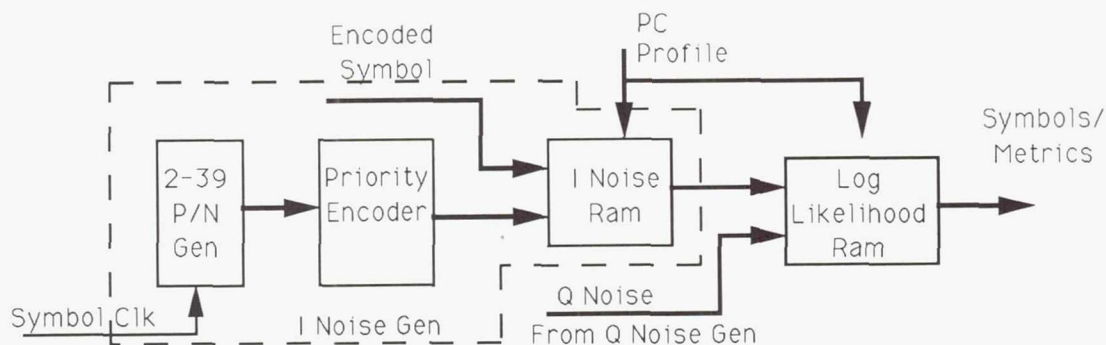
Fig. 5 Noise Generator / Log Likelihood

HARDWARE DESIGNS

Both the FHSC and the TDE chassis are fabricated using off the shelf 19" rack mountable chassis. They each contain their own power supplies. The circuits are designed onto MuPac multi-technology wire-wrap cards which fit into a MuPac VME type card cage. There are four card types in each of the chassis. One of the cards in the TDE is a PC design. This

design uses the ECLIPS logic family and clocks to rates in excess of 300 MHz. The wire wrap cards have been auto wrapped from files extracted from CAD captured schematics. All high speed signal interfaces are differential ECL. With the exception of all clocks the signals interface through D connectors on the front and back panels.

The ASIC is designed and scheduled for delivery in early August. The design was fully simulated using VLSI logic simulator tools. The simulations were conducted over the full commercial temperature. Results of these simulations indicate the ASIC will fully support the FHSC 300 megabit requirement over this temperature range. These simulations include the complete encoding and decoding of corrupted random data.

Simulations of the FHSC formatting circuits and the Chase pre and post-processors are currently being run. These simulations are at a card level. Completion of this effort is anticipated in early August.

OBSERVATIONS

A high-speed, high-rate coding technique suitable for both burst and continuous systems has been presented. It can operate as a single chip hard decision codec or, with the decoding applique, can utilize soft decision information in the decoding process. Coding gains up to 4 dB are obtained in the hard decision mode, increasing to 5.5 dB with soft decisions (at 10-8 BER).

Error correction coding has long been considered a good means to lower the required EIRP in communication systems having unlimited bandwith. However, high-rate codes such as the one described are also well suited for bandwith efficient systems. The CODEC rate and interface are matched to the larger signaling alphabets used for constrained bandwidth communications. Performance data indicates that coding gain improves slightly with increasing modulation alphabet size and is a week function of code word length. Even with the overhead required to insert parity bits, the net result is less power required to communicate a given data rate (say 300 Mb/s) over a fixed bandwidth channel (say 200 Mb/s).

The approach is extremely flexible by design. Hard and soft decision operation supports several different interface modes at data rates up to 300 Mb/s. Soft decision operation increases performance by as much as 1.5 dB. The soft decision applique can be easily translated into an ECL gate array considerably reducing the power and size of the FHSC.

It is believed that the approach and hardware resulting from this project will prove useful to a variety of high rate systems.

References

1) Polkinghorn, F. Jr., "Decoding of Double and Triple Error Correcting Bose-Chaudhuri Codes," IEEE Transactions of Information Theory, October 1966. pp. 480-481.

2) Chase, D., "A Class of Algorithms for Decoding Block Codes With Channel Measurement Information," IEEE Transactions on Information Theory, January 1972, pp.170-182.

3) Boyd, R. and Hartman, W., "Flexible High Speed CODEC," Advanced Modulation and Coding Technology Conference Publication, June 1989, Session 3, article 6.

Page intentionally left blank

PROGRAMMABLE DIGITAL MODEM[*]

John J. Poklemba
COMSAT Laboratories
Clarksburg, Maryland 20871

INTRODUCTION

In this paper the design of the Programmable Digital Modem (PDM) will be outlined. The PDM will be capable of operating with numerous modulation techniques including: 2-, 4-, 8-, and 16-ary phase shift keying (PSK), minimum shift keying (MSK), and 16-ary quadrature amplitude modulation (QAM), with spectral occupancy from 1.2x to 2x the data symbol rate. It will also be programmable for transmission rates ranging from 2.34 to 300 Mbit/s, where the maximum symbol rate is 75 Msymbol/s. Furthermore, these parameters will be executable in independent burst, dependent burst, or continuous mode. In dependent burst mode the carrier and clock oscillator sources are common from burst to burst.

To achieve as broad a set of requirements as these, it is clear that the essential signal processing must be digital. In addition, to avoid hardware changes when the operational parameters are changed, a fixed interface to an analog intermediate frequency (IF) is necessary for transmission.; and, common system level architectures are necessary for the modulator and demodulator. Lastly, to minimize size and power as much of the design as possible will be implemented with application specific integrated circuit (ASIC) chips.

MODULATOR ARCHITECTURE AND DESIGN

Baseband vs IF Digital-to-Analog Sample Conversion

Should the modulator output analog samples at baseband or IF? To answer this, the restrictions caused by the digital-to-analog (D/A) conversion device will first be examined. A D/A converter is inherently a sample-and-hold device that imposes a lowpass $\sin(x)/(x)$ envelope on the baseband output spectrum and its replicas. This effect is shown in Figure 1a for the integer minimum Nyquist sample rate of two samples/symbol (s/s) and square root 40-percent raised cosine spectral shaping. To support most of the two-dimensional modulation formats listed above, four complex s/s or equivalently two in-phase and two quadrature channel s/s are required. The gap between the main lobe and the first replicated spectra allows a practical analog reconstruction filter to be used, and the D/A stopband notches provide inherent filtering as they occur in the center of the replicated spectra.

To convert the digital baseband samples directly to an IF output at a minimum number of s/s implies that their spectra be shifted up in frequency. To avoid restricting the upper data rate of operation, 3 s/s is the minimum that can be used for IF sampling as shown in Figure 1b. Because of the spectral shift, the D/A converter would cause a considerable amount of amplitude skew across the IF passband; and the first replicated image, centered just above $2R_s$, is very close to the desired lobe, centered just below R_s.
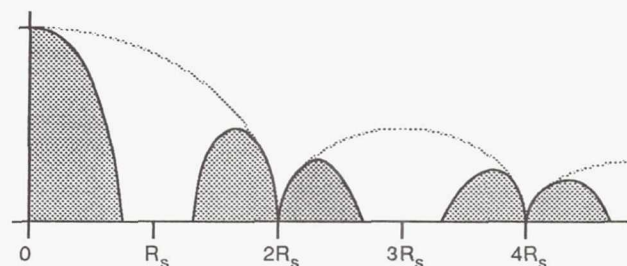
a. Baseband Sample Conversion at 2 Samples/Symbol
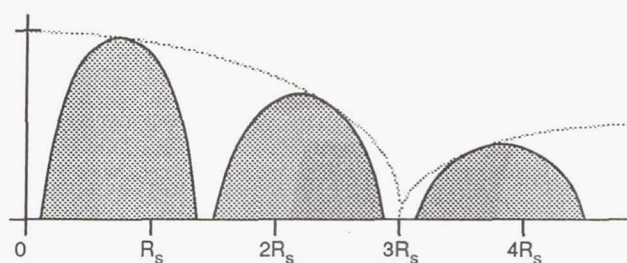
b. IF Sample Conversion at 3 Samples/Symbol

Figure 1. D/A Aperture Effects

So even at the minimum bandpass sample rate, it is very difficult to filter out the replicated spectra. Hence, it's clear that for a given speed capability in the digital hardware, baseband sampling will achieve higher data rate operation. Thus, at such high speeds, the most effective way to process the data is with a minimum integer number of samples per symbol with parallel in-phase and quadrature (I and Q) channels at baseband, and analog quadrature carrier mixing for conversion to an IF.

To accommodate multirate operation, the sample rate into the D/A converter will always be within the octave range of 75-150 Msample/s, regardless of the data rate; and the number of samples per symbol will always be a power of two. In this manner, the sample clock replicated spectra of Figure 1a can be removed over the entire symbol rate range of operation with a single analog reconstruction filter. Moreover, the highest symbol rate range is 37.5-75 Msymbol/s at two s/s. The next octave range down is then 18.75-37.5 Msymbol/s at four s/s, and so on.

The replication removal filter must pass as much of the main lobe at the maximum symbol rate (R_S = 75 Msymbol/s) as possible, while rejecting the low end of the first replicated lobe at a symbol rate an octave below the maximum (R_S = 37.5 Msymbol/s). A good compromise, determined in conjunction with the bit error rate (BER) simulations, is an

elliptic lowpass filter with a 0.2 dB equiripple passband extending from DC to 48 MHz, with a stopband beginning at 64 MHz of minimum attenuation greater than 30 dB. The sample-and-hold effect of the D/A provides additional filtering to suppress the sample clock replications below 40 dB. To avoid additional analog hardware, group delay dispersion in the replication removal filter will be compensated with digital processing.

A block diagram of the basic modulator architecture is given in Figure 2. The modulator is divided into a digital baseband processor with an analog quadrature carrier IF. The primary function of the baseband processor is to spectrally shape or filter the data in a bandwidth efficient manner, and to convert it to a baseband quadrature format prior to carrier modulation. The quadrature format supports nearly any modulation format that can be represented in a two-dimensional signal space, and the parallel I and Q channels support higher rate operation. The analog portion of the modulator then performs the function of translating the I and Q data representation on to cosine and sine carriers, respectively.

Transmit Spectral Shaping

To achieve the best BER performance possible, it would be desirable to digitally implement and match the transmit and receive filter spectra with a square root Nyquist characteristic, assuming that the remaining filtering functions in the transmission link are transparent. However, in general, the transmit and receive data filters cannot be matched and must be predistorted to account for replication removal, IF, and anti-aliasing filters as well as transmission link impairments.

Because of the strict magnitude and phase constraints for Nyquist data filters, the most appropriate digital filter implementation is the finite impulse response (FIR), which inherently has linear phase. A greatly simplified equivalent implementation is possible because the transmit symbols have relatively few deterministic levels; i.e., BPSK, QPSK,

and MSK only require two input levels. The reduced complexity implementation involves a memory table lookup. A brief description is as follows. Input data symbols are read into a shift register whose length is equal to the number of symbols in the impulse response aperture to be represented. To determine the transmit impulse response, all of the link frequency responses are cascaded, and a discrete Fourier transform (DFT) is employed to compute the predistorted samples. A fast Fourier transform (FFT) is not used because, in general, the sample sets are not a power of N. The symbol patterns in the shift register change every symbol time, so for each symbol pattern there is a unique set of precomputed sample values that will be clocked out of the memory. That is, within a given symbol pattern, there are N unique samples per symbol. The memory size required is determined from

$$M^L \bullet N \tag{1}$$

where

M	=	number of in-phase or quadrature symbol amplitude levels required
L	=	length of the filtering aperture in symbol times
N	=	number of samples per symbol.

Hence, the memory size increases linearly with the number of s/s, but geometrically versus impulse response aperture length and the number of I or Q amplitude levels. For example, a 16-PSK signal constellation will be represented with eight I/Q levels (±4); whereas QPSK requires only two I/Q levels. Several permutations of the maximum memory sizes required are listed in Table 1 for 32 s/s. The common achievable size for all of the modulation techniques is indicated in parentheses, 131K bytes. Approximate carrier spacings that may be supported are also listed.

The best combination of high density and speed memory currently available is 65K x 4 with an access time of 8 ns, which when setup, hold, and skew times are included, provides a small amount of timing margin for operation at
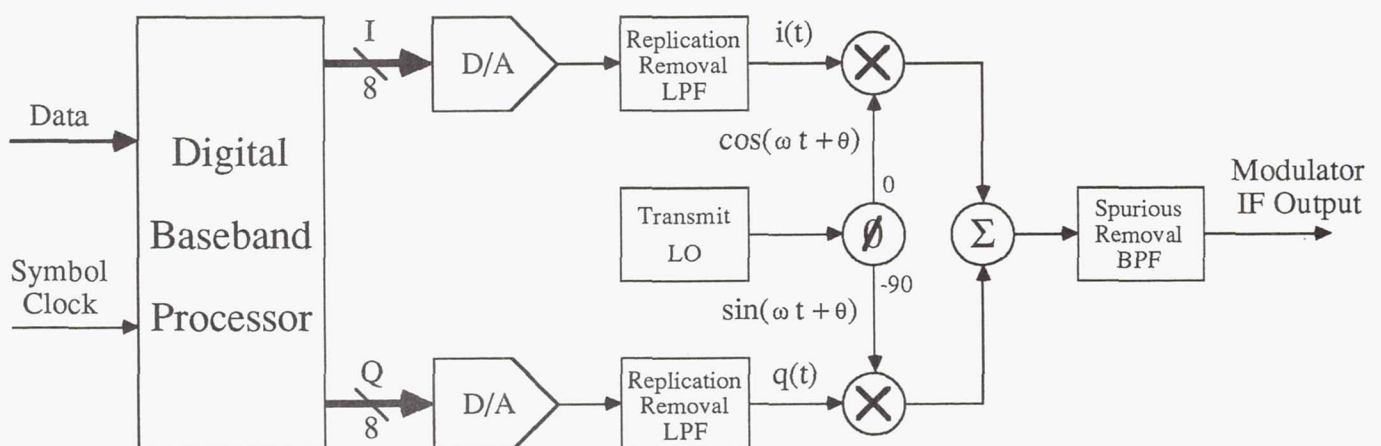
Figure 2. Basic Modulator Architecture

264

Table 1. Maximum I or Q Channel Memory Requirements at 32 Samples/Symbol

MODULA-TION TECHNIQUE	NUMBER OF SIGNAL LEVELS	APERTURE LENGTH (SYMBOLS)					
		3	4	5	6	8	12
BPSK, MSK, QPSK	2(±1)	256	512	1k	2k	8.2k	(131k)
8-PSK, 16-QAM	4(±X,±Y)	2k	8k	33k	(131k)	2.1M	
16-PSK	8 (±A,±D; ±B,±C)	16.4k	(131k)	1.0M	8.4M		
Carrier Spacing (R_S Multiples)		1.9	1.8	1.6	1.4	1.3	1.2

75 Msymbol/s (13.3 ns). For 8-bit resolution, four of these chips are required in each of the I and Q channels, along with the 12-symbol shift register. This is considerably simpler than an equivalent 384-tap FIR filter implementation with its incumbent set of digital multiplies and sums. The 8-bit output resolution for the memory results in good spectral quantization noise, which is >40 dB down over the range of rates desired.

DEMODULATOR ARCHITECTURE AND DESIGN

IF vs Baseband Analog-to-Digital Sample Conversion

The issue of sampling directly at IF vs conversion to baseband prior to sampling will now be analyzed separately for the demodulator. With IF sampling, the IF center frequency will scale with the data rate unless a noninteger number of samples per symbol or more complex processing is used. To handle a noninteger number of samples per symbol, an interpolating filter is needed. In the demodulator, the interpolating filter would basically perform two functions. It converts asynchronous samples to synchronous samples at two samples per symbol; such that over each symbol interval, one of the samples occurs at the data detection sample point, while the other is at the average value of the zero crossings for symbol timing recovery. However, an interpolating filter is hardware intensive and speed restrictive. Furthermore, to operate at 75 Msymbol/s suggests that the lowest IF center frequency be at least 75 MHz, or more suitably 140 MHz. A half-cycle of the carrier sinusoid at this rate is about 3.5 ns. The narrowest sampling aperture on currently available analog-to-digital (A/D) converters is on the order of 1.5 to 2 ns. Hence, the width of the sampling aperture is approximately one-half of the slowest practical positive or negative carrier excursion. This imposes a lowpass sin (x)/(x) envelope on the incoming bandpass spectra, as was illustrated in Figure 1. For a 1.75-ns aperture, the sin (x)/(x) envelope is about 1 dB down at 140 MHz, so sampling at IF would also cause a variable amplitude skew across the passband for the higher operational data rates. As a result of limitations due to the A/D sampling aperture and interpolating filter realizations, the receive bandpass signal will be down converted with

carriers in phase quadrature for subsequent sampling at baseband.

The requirements for the anti-aliasing filter to limit the incoming bandwidth prior to A/D conversion are very comparable to those for the replication removal filter in the modulator. For example, the bulk of the main spectral lobe must be passed at the maximum symbol rate, which extends from DC to 52.5 MHz for a 40-percent Nyquist channel. In addition, the filter must restrict the incoming noise bandwidth to half the minimum sample rate to avoid aliasing at the higher data rates. For this and other reasons which will be explained subsequently, the minimum sample rate on the demodulator, 100 Msample/s is higher than that on the modulator, 75 Msample/s. Previous simulations have shown that 30 dB stopband attenuation is sufficient to have negligible impact on BER, and that greater attenuation merely makes it more difficult to compensate for the filter's delay dispersion. Hence, for simplicity, the anti-aliasing filter will be designed with identical parameters as the modulator replication removal filter. This also allows for a common IF hybrid or MMIC to be developed for use in both the modulator and demodulator.

Demodulator Block Diagram

The basic demodulator structure is given in Figure 3. Note the interdependence of the acquisition estimate processor, the data detection, and the recovery loops. A GaAs ASIC chip is currently being developed that will contain two programmable MACs. It will be capable of being reconfigured to operate in nine separate locations in the demodulator. The ASIC multipliers will be 8 x 8 with a 16-bit barrel shifted output, and the accumulators will be 24 bits with 16-bit preloading. All of the required ASICs will be capable of 150-Msample/s pipeline operation.

Receive Data Detection Filter

The most potentially hardware-intensive function in the demodulator is the receive data detection filter. A memory-based structure is not feasible because of the large number of input quantization levels due to channel impairments and noise. A minimum complexity FIR filter with a reduced or decimated output sample rate is desired. This can be achieved with a very high-speed multiplier-accumulator (MAC), where each accumulator output sample corresponds to a weighted average of a set of incoming samples. Since the output of this filter will feed all of the remaining processing stages necessary in the demodulator, it has been dubbed the "pre-averager" data filter. Separate even and odd MACs are required because the input sample sets that the pre-averager must process are overlapping, as shown in Figure 4 [1]. The even samples are used for data detection, carrier recovery, and gain control; whereas the odd samples provide symbol timing recovery. As indicated, the averages are taken over N samples in one-symbol intervals. So, in effect, the pre-averager impulse response extends over a one-symbol aperture. However, BER simulations with adjacent channels on 1.4x the symbol rate spacings have shown that a one-symbol aperture is not adequate, regardless of the weighting function employed. What is necessary is a
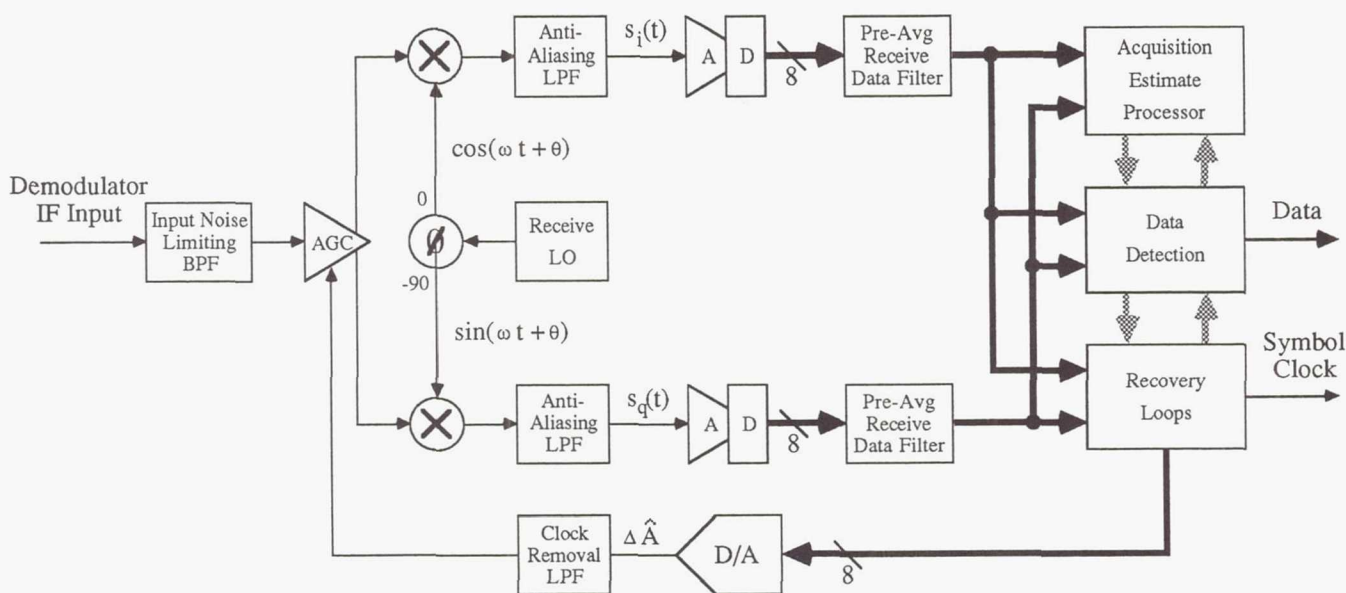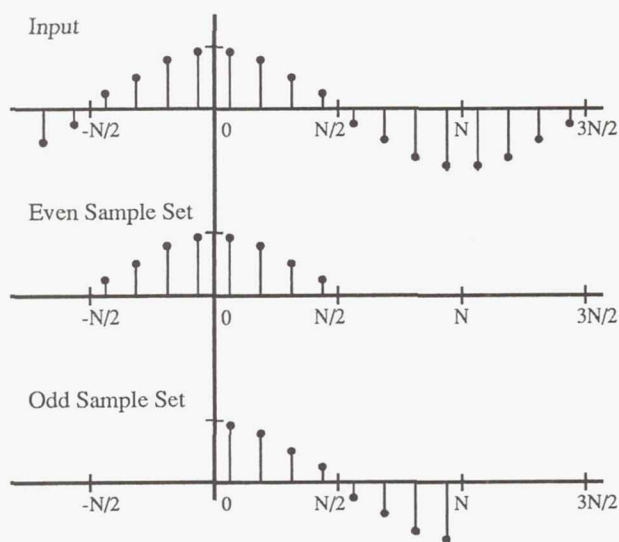
Figure 3. Basic Demodulator Architecture

Figure 4. Overlapping Pre-Averaged Sample Sets

sharper rolloff filtering function that has a stopband in the region above 0.7 R_S (half the center-to-center carrier spacing) to remove adjacent channel interference and noise.

Receive Data Filter Impulse Response Derivation

From an implementation point of view, the most straightforward way to modify the poor adjacent channel rejection (ACR) capability of the one-symbol aperture pre-averager is to increase its aperture to two symbols, with 50 percent overlapping averaging intervals. Next, it would be

desirable to find a strictly time-limited two-symbol-long impulse response, with a stopband above 0.7 R_S. Proceeding to the sampled frequency domain, a very general Nyquist filtering function may be defined to satisfy this condition for two s/s as follows

$$
\begin{aligned}
H(0) &= 1.0 \\
H(1) &= 0.5 \\
H(2) &= 0.0 \\
H(3) &= 0.5
\end{aligned}
\tag{2}
$$

where R_S has been normalized to 2.

These four frequency domain samples at two s/s will yield four time domain samples that extend over a two-symbol aperture. Using the definition of the inverse DFT,

$$
h(n) = \frac{1}{N} \sum_{k=0}^{N-1} H(k) \exp\left(j\, 2\pi k n / N\right), \quad 0 \le N \le N\text{-}1
\tag{3a}
$$

on the values in equation (2) yields a raised cosine pulse:

$$
h(n) = \frac{1}{4}\left\{1 + 0.5\left[\exp\left(j\pi n / 2\right) + \exp\left(j 3\pi n / 2\right)\right]\right\}
\tag{3b}
$$

$$
= \frac{1}{2}\left[\frac{1 + \cos\left(\pi n / 2\right)\exp\left(-j\pi n\right)}{2}\right]
\tag{3c}
$$

$$
= \frac{1}{2}\left[\frac{1 + \cos\left(\pi n / 2\right)}{2}\right]
\tag{3d}
$$

where the exponential phase term is dropped from the last equality because the cosine term is zero for n-odd, and it has no effect for n-even.

266

Extensive BER simulations have shown the raised cosine pulse (RCP) impulse response of equation (3d) to be substantially more effective than truncated square root Nyquist impulse responses in providing good adjacent channel rejection, for a two-symbol aperture filter at any number of samples per symbol. However, using the RCP response implies that the bulk of the Nyquist channel characteristic resides in the demodulator, so matched filtering has been sacrificed for a simplified implementation that is effective in rejecting adjacent channels. Simulations have shown that this transmit/receive filter apportionment causes a degradation on the order of 0.5 dB in BER.

The frequency responses for the raised cosine pulse at 2, 3, 4, and 32 s/s are depicted in Figures 5a, b, c, and d, respectively. Observe that the ACR improves as the number of s/s is increased. Fortunately, at two s/s the analog anti-aliasing filter provides most of the needed ACR. Moreover, it is necessary to include additional integer sample rates in the demodulator between 2, 4, and 8 s/s, namely, 3 and 6 s/s to provide sufficient ACR. The relationship between sample and symbol rates as well as the number of s/s in the modulator and demodulator are listed in Tables 2a and 2b, respectively.

Table 2a. Modulator Rate Ranges
(Msymbol/s, Msample/s)

SYMBOL RATE	SAMPLES/SYMBOL	SAMPLE RATE
2.34375–4.6875	32	75–150
4.6875–9.375	16	75–150
9.375–18.75	8	75–150
18.75–37.5	4	75–150
37.5–75.0	2	75–150

Table 2b. Demodulator Rate Ranges
(Msymbol/s, Msample/s)

SYMBOL RATE	SAMPLES/SYMBOL	SAMPLE RATE
2.34375–4.6875	32	75–150
4.6875–6.25	24	112.5–150
6.25–9.375	16	100–150
9.375–12.5	12	112.5–150
12.5–18.75	8	100–150
18.75–25.0	6	112.5–150
25.0–37.5	4	100–150
37.5–50.0	3	112.5–150
50.0–75.0	2	100–150

To summarize, the pre-averager has several significant properties: 1) it serves as a variable rate FIR receive data filter of minimal complexity; 2) it reduces the processing rate and complexity of subsequent circuitry to 1 s/s; 3) it reduces the incoming noise bandwidth to approximately ±Rs/2, thereby improving the input signal-to-noise (S/N) ratio established by the fixed analog anti-aliasing filter.

Data Detection
Data detection for the various modulation techniques is achieved with a memory table lookup of the even samples from the (I, Q) signal vector out of the pre-averagers. The sampling is synchronous and the symbol timing recovery loop will cause the even samples to automatically occur at the optimum data detection time instant. As stated previously, the largest memory size available at 75-MHz signaling speeds is 64K x 4, which provides for an I and Q input resolution of 8 bits.

Steady-State Recovery Loop Architecture
In 1977, a joint estimator-detector approach was developed at COMSAT Laboratories to provide an optimum way to recover carrier and clock for QPSK data transmission. It was found that the resultant technique which was dubbed Concurrent Carrier and Clock Synchronization (CCCS) applies to many types of digital data modulation. In particular, the CCCS technique is applicable to any modulation format that can be represented in quadrature carrier form: such as BPSK, QPSK, ... M-ary PSK, QAM, MSK, etc. Hence, this technique provides a basis for the PDM demodulator structure. Details of the CCCS technique are contained in References 2 and 3.

Some of the salient CCCS features which impact the PDM architecture will now be discussed. The CCCS method demonstrated that the optimum steady-state carrier phase and clock timing estimators are phase-locked loops (PLLs), which use post-detection feedback to remove data pattern noise and generate error signals that drive the loops. Post-detection data feedback is essentially noiseless because, even at a relatively poor BER of 10^{-2}, only 1 of every 100 detected data bits is incorrect. Hence, the loop S/N is merely reduced by a factor of 0.98 (-0.09 dB). Apart from knowing the transmitted data sequence, this is as well as a recovery loop can do.

For more complex signaling formats such as 8-, 16-PSK, and 16-QAM, where a quadrature carrier description of the IF signal requires several amplitude levels to be represented, the CCCS detected data feedback in the recovery loops must be multilevel. Multilevel feedback gives the larger average S/N samples proportionally more weight than the smaller ones, thereby maintaining the optimality of the recovery loop S/Ns. Moreover, the CCCS approach enables a common carrier, clock, and gain control recovery loop architecture to be used for any modulation format that can be represented in quadrature carrier form.

The basic error signal mechanism and loop filter for tracking in the CCCS architecture is illustrated in Figure 6. Table 3 lists the feedback signals needed for automatic gain control (AGC), carrier, and clock tracking. This common structure can be reconfigured in a MAC format by performing the multiplications sequentially and summing their products. Although this doubles the maximum speed requirement from 75 to 150 Msample/s, it is consistent with the speed already necessary for the pre-averager.

The error signals that drive the tracking loops are each processed by a loop filter to provide an output estimate.
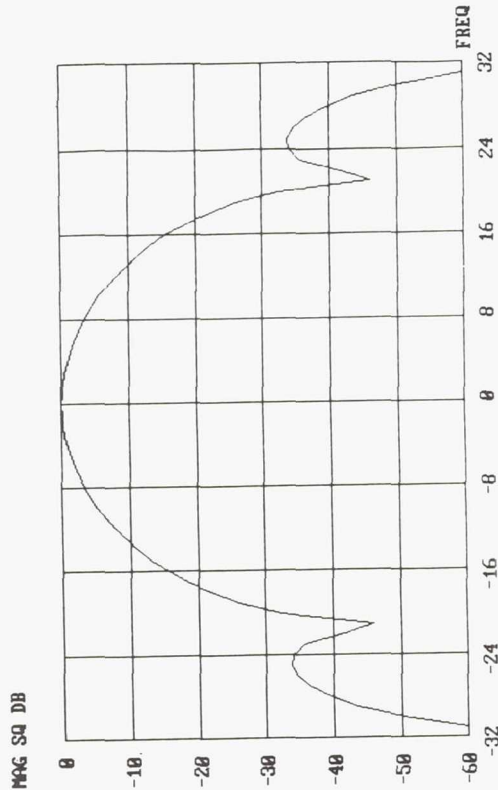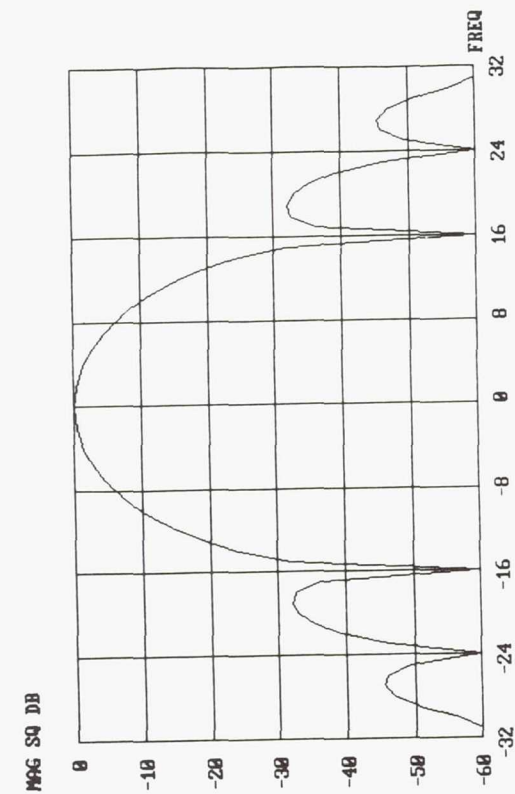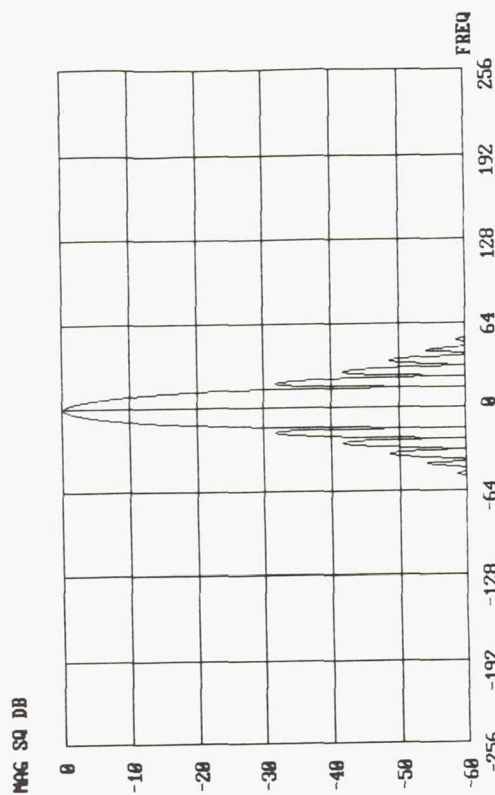
a. 2 Samples/Symbol

b. 3 Samples/Symbol

c. 4 Samples/Symbol

d. 32 Samples/Symbol

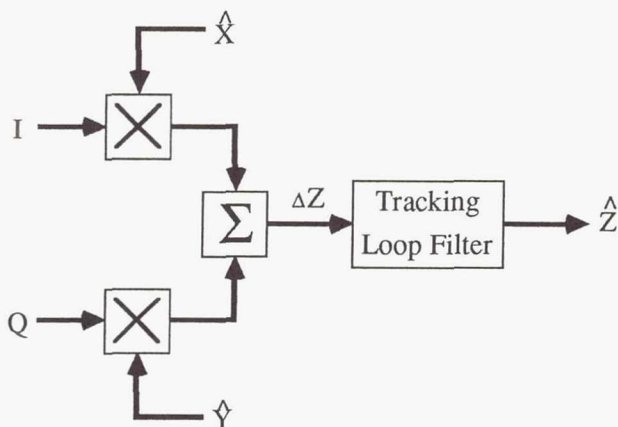*Figure 5. Raised Cosine Pulse Filter Frequency Response*

268

Figure 6. Recovery Loop Processor

Previous experience has shown that the AGC and clock loops need only be first order, whereas to track frequency offsets, the carrier loop must be second order. Hence, MACs will also be employed to satisfy the loop filter requirements. The loop bandwidth parameters can then be programmed by changing the multiplier gain constant. Moreover, the MAC architecture can be applied at numerous locations in the demodulator, including the acquisition circuitry.

Table 3. Tracking Loop Error Feedback Signals

FUNCTION	$\hat{X}$	$\hat{Y}$	$\hat{Z}$
Amplitude Level	$\hat{I}$	$\hat{Q}$	$\Delta\hat{A}$
Carrier Phase	$\hat{Q}$	$-\hat{I}$	$\hat{\theta}$
Symbol Timing	$\Delta\hat{I}$	$\Delta\hat{Q}$	$\hat{\tau}$

Notes: $\Delta\hat{A} = \hat{A} - A_{ref}$

$\Delta\hat{I} = \hat{I}(kT_s) - \hat{I}[(k-1)T_s]$

$\Delta\hat{Q} = \hat{Q}(kT_s) - \hat{Q}[(k-1)T_s]$

The output of the first order AGC loop filter is the estimate of the amplitude level error, $\Delta\hat{A}$; which is the control signal for the AGC amplifier. The AGC amplifier gain, G, is modeled as

$$G = \frac{G_{nom}}{1 + \Delta\hat{A}/A_{ref}}, \quad \Delta\hat{A} > -A_{ref} \qquad (4)$$

where G_{nom} is the nominal gain when $\Delta\hat{A} = 0$.

The output of the second order carrier loop filter is the estimate of the phase of the incoming signal. It includes the linear phase variations modulo 180° necessary to track carrier frequency offsets. Since a fixed frequency local oscillator (LO) is employed to down-convert the incoming signal to baseband, a carrier beat frequency occurs in the demodulated I and Q channels. A carrier phase rotator is used to eliminate the beat after the pre-averager data filters,

prior to detection. If the generalized incoming QAM signal is defined as

$$s[t, A, \theta(t), \tau] \triangleq A\{i(t,\tau) \cos[\omega t + \theta(t)] + q(t,\tau) \sin[\omega t + \theta(t)]\} \quad (5a)$$

where

A	=	incoming signal amplitude
ω	=	incoming signal frequency
$\theta(t)$	=	incoming signal phase uncertainty
$i(t,\tau)$	=	filtered in-phase modulating waveform
$q(t,\tau)$	=	filtered quadrature modulating waveform
τ	=	modulating waveform timing uncertainty

and the quadrature LO outputs for down-conversions are

$$lo_i(t) = 2 \cos(\omega t) \qquad (5b)$$

$$lo_q(t) = 2 \sin(\omega t) \qquad (5c)$$

The resulting baseband I and Q components prior to phase rotation are then

$$s_i(t) = A\{i(t,\tau)\cos[\theta(t)] + q(t,\tau)\sin[\theta(t)]\} \qquad (6a)$$

$$s_q(t) = A\{q(t,\tau)\cos[\theta(t)] - i(t,\tau)\sin[\theta(t)]\} \qquad (6b)$$

To decouple the I and Q modulating waveforms, the carrier phase rotation is defined as

$$\begin{bmatrix} s_i'(t) \\ s_q'(t) \end{bmatrix} = \begin{bmatrix} \cos[\hat{\theta}(t)] & -\sin[\hat{\theta}(t)] \\ \sin[\hat{\theta}(t)] & \cos[\hat{\theta}(t)] \end{bmatrix} \begin{bmatrix} s_i(t) \\ s_q(t) \end{bmatrix} \qquad (7a)$$

$$\begin{bmatrix} s_i'(t) \\ s_q'(t) \end{bmatrix} = A \begin{bmatrix} i(t,\tau)\cos[\Delta\theta(t)] + q(t,\tau)\sin[\Delta\theta(t)] \\ q(t,\tau)\cos[\Delta\theta(t)] - i(t,\tau)\sin[\Delta\theta(t)] \end{bmatrix} \qquad (7b)$$

where $\Delta\theta(t) = \theta(t) - \hat{\theta}(t)$, and the output estimate from the carrier loop filter is converted into two quadrature cosine and sine terms. The phase rotation described in equations (7) will also be implemented with MACs.

In the symbol timing tracking loop, the first order loop filter is actually a numerically controlled oscillator (NCO); which has an accumulator that holds the timing phase. Hence, the error signal from the timing phase detector is added with appropriate weighting to a constant that sets the nominal sample clock frequency, NR_s at the NCO input. The symbol clock as well as all other clocks used in the demodulator are then synchronously divided down from NR_s.

Burst-Mode Synchronization Techniques

To expedite lock and provide a high degree of false-and-miss detection reliability in burst mode, a parallel acquisition estimate path has been added to the tracking loop architecture. The initial carrier and clock phase as well as the amplitude level are estimated in this path and injected directly into the recovery loop accumulators. This effectively minimizes the loop lock-up transients. Since the accuracy of the acquisition measurement is proportional to the length of its observation interval, the burst false-and-miss detection probabilities can be made arbitrarily small.

In computing the acquisition estimates, it is desirable to uncouple them so they may be processed independently,

269

thereby having fewer degrees of uncertainty. For modulation techniques whose I and Q channels are not time staggered (such as offset formats), independent parallel processing of the estimates is possible with "01" modulation in both channels [4],[5]. The analog baseband I and Q signals defined in equations (6a and b) then may be described by

$$s_i(t) = \sqrt{2} \; A \sin[\pi R_s(t+\tau)] \; \{\cos[\theta(t)] + \sin[\theta(t)]\} \quad (8a)$$

$$s_q(t) = \sqrt{2} \; A \sin[\pi R_s(t+\tau)] \; \{\cos[\theta(t)] - \sin[\theta(t)]\} \quad (8b)$$

Equations (8a and b) can be reduced to

$$s_i(t) = 2A \sin[\pi R_s(t+\tau)] \; \sin[\theta(t) + \pi/4] \quad (9a)$$

$$s_q(t) = 2A \sin[\pi R_s(t+\tau)] \; \cos[\theta(t) + \pi/4] \quad (9b)$$

In the sampled domain, equations (9a and b) are rewritten as

$$I_{2k} \triangleq 2A(-1)^k \cos(\phi_\tau/2) \sin(\theta + \pi/4) \quad (10a)$$

$$Q_{2k} \triangleq 2A(-1)^k \cos(\phi_\tau/2) \cos(\theta + Pi/4) \quad (10b)$$

$$I_{2k-1} \triangleq 2A(-1)^k \sin(\phi_\tau/2) \sin(\theta + \pi/4) \quad (10c)$$

$$Q_{2k-1} \triangleq 2A(-1)^k \sin(\phi_\tau/2) \cos(\theta + \pi/4) \quad (10d)$$

where the timing phase offset, $\phi_\tau = 2\pi R_s\tau$, and the subscripts 2k and 2k-1 denote even and odd samples, of the kth symbol, respectively.

Amplitude Level Acquisition Estimate

The most straightforward way to extract the amplitude A from equations (10a through d) independent of the phase and timing uncertainties is squaring, and then averaging to improve the estimate SNR. To simplify the hardware implementation and allow for sharing of common processing elements, the averaging should be done as soon as possible to lower the output sample rate. Because of the carrier frequency offset, the even and odd pairs of samples must be squared and combined in MACs on a symbol-by-symbol basis and then averaged.

$$E^2 \triangleq \sum_k \left(I_{2k}^2 + Q_{2k}^2\right) = 4A^2 \cos^2(\phi_\tau/2) \quad (11a)$$

$$O^2 \triangleq \sum_k \left(I_{2k-1}^2 + Q_{2k-1}^2\right) = 4A^2 \sin^2(\phi_\tau/2) \quad (11b)$$

Equations (11a and b) can then be combined to give the amplitude level estimate

$$\widehat{A} = \sqrt{E^2 + O^2}/2 \quad (12)$$

Equation 12 is most easily implemented as a memory table lookup. It was found in the emulations that 10 bits of resolution are needed for E^2 and O^2 because of the squaring. An intermediate compression table lookup is necessary to reduce the memory size in implementing equation (12) from 1 Mbyte to 64 kbytes.

Carrier Phase Acquisition Estimate

In reviewing equations (10a through d), it is apparent that there are several ways to isolate the carrier phase offset. For instance, the phase can be computed on a symbol-by-symbol basis as the arctangent of linear, square, or absolute value functions of I/Q, and then averaged; or I and Q can be squared first, and then averaged and processed as the arctangent of the sum of squares; or I and Q may be premultiplied by the preamble to remove the modulation, averaged, and the arctangent taken. All of these techniques have relative advantages and disadvantages. For instance, squaring the incoming samples increases the twofold ambiguity with "01" preamble modulation to fourfold; which either increases the complexity of the unique word detector or requires additional acquisition processing to unravel. Computing the arctangent on a symbol-by-symbol basis does not allow the arctangent processing element to be shared with the symbol timing loop. So the method chosen is the latter of the three examples for the following reasons. Premultiplication of the incoming samples by the known preamble removes the data modulation without S/N degradation. By next averaging the samples prior to the nonlinear arctangent operation, the S/N is improved. Finally, the largest pair of odd or even sample sums are chosen for the arctangent, so the twofold phase ambiguity is maintained. To make the odd vs even decision, the O^2 and E^2 sums, which were previously calculated in the amplitude level estimator are compared. Hence the resulting carrier phase estimate is computed from the ratio of I over Q samples as

$$\widehat{\theta} = \tan^{-1}\left(\frac{\pm \Sigma|I_{2k} \text{ or } I_{2k-1}|}{\pm \Sigma|Q_{2k} \text{ or } Q_{2k-1}|}\right) - \pi/4 \quad (13)$$

where

$$E^2 \underset{2k-1}{\overset{2k}{\gtrless}} O^2$$

Equation (13) will be implemented as a 64-kbyte memory table lookup.

To find the frequency offset, two such phase estimates are computed over the first and second halves of the preamble as θ_1 and θ_2, respectively. The frequency offset can then be computed from the phase difference as

$$\Delta\omega = \frac{\Delta\theta}{\Delta T} = \frac{\theta_2 - \theta_1}{P/2} \quad (14)$$

where P is the total length of the preamble in symbol time units. The end-of-preamble phase estimate is determined from the measured phase and frequency difference as

$$\theta_{EOP} = \theta_2 + \Delta\omega \cdot \Delta T \quad (15)$$

Equations (14) and (15) will also be implemented as 64-kbyte memory table lookups.

Symbol Timing Acquisition Estimate

Again, there are several ways to compute the initial symbol timing error. It could be calculated from the arctangent of the square root of the previously computed values O^2/E^2, but the squaring would cause a twofold timing ambiguity which requires additional processing to resolve. It can also be computed from the arctangent of the largest pair of preamble premultiplied odd and even samples, which also requires an I^2 or Q^2 largest decision. The latter case turns out to be easier to implement since two of the tracking loop MACs are idle during acquisition and can be employed to calculate I^2 and Q^2; and in addition, the arctangent operation can be time shared with that required for carrier phase acquisition. So the symbol timing offset is computed from the ratio of odd over even samples as

$$\phi_\tau = 2 \tan^{-1}\left(\frac{\pm \Sigma |I_{2k-1} \text{ or } Q_{2k-1}|}{\pm \Sigma |I_{2k} \text{ or } Q_{2k}|}\right) \qquad (16)$$

where

$$I^2 \triangleq \sum_k \left(I_{2k}^2 + I_{2k-1}^2\right) \underset{Q}{\overset{I}{\gtrless}} Q^2 \triangleq \sum_k \left(Q_{2k}^2 + Q_{2k-1}^2\right)$$

Equation (16) will share the same 64-kbyte memory table as the carrier phase in equation (13). The slight differences in the expressions will be compensated for in the end of preamble phase computation from equations (14) and (15).

CONCLUSIONS

Operation of digital signal processing (DSP) circuitry at sample rates as high as 150 MHz appears feasible. The two most speed-critical areas are memories and multiplier-accumulators. Currently available high-density static RAMs can only operate up to approximately 80 MHz and must be ping-ponged to achieve the desired rate. The workhorse of the processing is clearly the multiplier-accumulator. To achieve 150-MHz operation with sufficient margin and power efficiency, GaAs is the most appropriate technology; potential GaAs vendors have recommended a standard-cell rather than a gate-array approach for this application.

Subsequent hardware emulations have verified the fundamental design approach presented in this paper, as well as the bit resolutions and aperture lengths used. The results will be submitted in a forthcoming publication.

REFERENCES

[1] J. J. Poklemba, C. Wolejsza, and J. Thomas, "Programmable Noise Bandwidth Reduction by Means of Digital Averaging," COMSAT Invention Disclosure No. 25-E-14, June 27, 1986.

[2] J. J. Poklemba, "Concurrent Carrier and Clock Synchronization for Data Transmission Systems," U.S. Patent No. 4, 419, 759, December 6, 1983.

[3] J. J. Poklemba, "A Joint Estimator-Detector for QPSK Data Transmission," COMSAT Technical Review, Vol. 14, No. 2, Fall 1984, pp. 211–259.

[4] S. A. Rhodes, "Synchronization for TDMA/QPSK Up-Link Transmissions With Digital On-Board Processing," COMSAT Technical Memorandum, CL-15-86, August 13, 1986.

[5] S. A. Rhodes, "Modification of Acquisition Algorithms for On-Board Demodulator," COMSAT Technical Note, CTD-87/007, January 14, 1987.

Page intentionally left blank

Multi-Rate Demodulator Architecture[*]

Michael A. Sherry and Gregory S. Caso
TRW Electronic Systems Group
Redondo Beach, CA 90278

I.   Summary

A unique digital Multi-Rate Demodulator (MRD) architecture is presented for onboard satellite communications processing. The multi-rate feature enables the same demodulator hardware to process either one wideband channel (WBC), or process up to thirty-two independent narrowband channels (NBC) that are time-division-multiplexed (TDM). The MRD can process many quadrature modulation format such as Offset-Quadrature-Phase-Shift-Keying (OQPSK). Possible applications include voice and data transmission for commercial or military users.

II.   Introduction

The MRD currently being developed, shown in Figure 1, is configured for Differentially Encoded OQPSK, with a WBC symbol rate of 1.024 Msymbol/sec. and NBC symbol rate of 32 Ksymbol/sec. OQPSK also referred to as Staggered-QPSK (SQPSK), since the Q baseband signal is staggered in time by 1/2 a symbol period relative to the I baseband signal. Each MRD can be configured for processing either a single high data-rate, WBC or for processing up to 32 low data-rate NBC's at 1/32 the WBC data rate. Reconfiguration from WBC to NBC's, or from NBC's to WBC, would be performed on-line. System parameters for the entire MRD were obtained through extensive Block-Oriented-Systems-Simulator (BOSS) simulation.
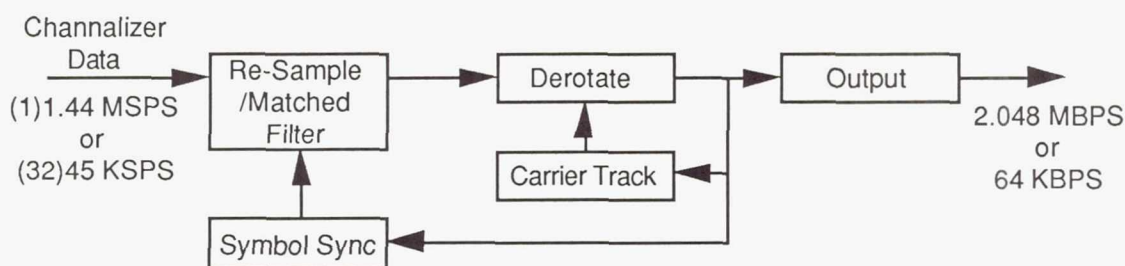
Figure 1. Multi-Rate Demodulator

Each MRD accepts as input, digitized, quadrature down-converted, baseband signals. Quadrature down-conversion decomposes the composite carrier waveform into in-phase (I) and quadrature (Q) components. Data is input to the MRD from a channelizer that performs the quadrature down-conversion, and multiplexing if NBC's. When the MRD is configured for WBC processing the input will be consecutive samples, at 1.44 Msample/sec., from a single channel. If the MRD is configured for NBC processing the input will be TDM'ed samples, at 1.44 Msample/sec., from 32 independent channels. The input data rate for a single NBC will effectively be 45 Ksample/sec. Input data ordering for WBC's and NBC's is shown in Figure 2. All demodulator processing for a channel, if NBC, or an input sample, if WBC, must be performed within the input sample period. If NBC's are processed the MRD hardware is time-shared, with each channel being allocated one input sample epoch to perform demodulator operations. The system clock frequency is 16 times the WBC input data rate with 16 system clock cycles being defined as an input sample epoch. Processing for any of the 32 NBC's is completely independent of the other channels.

WBCDn = Wideband Channel; Data Sample #n
NBCmDn = Narrow band Channel #m; Data Sample #n
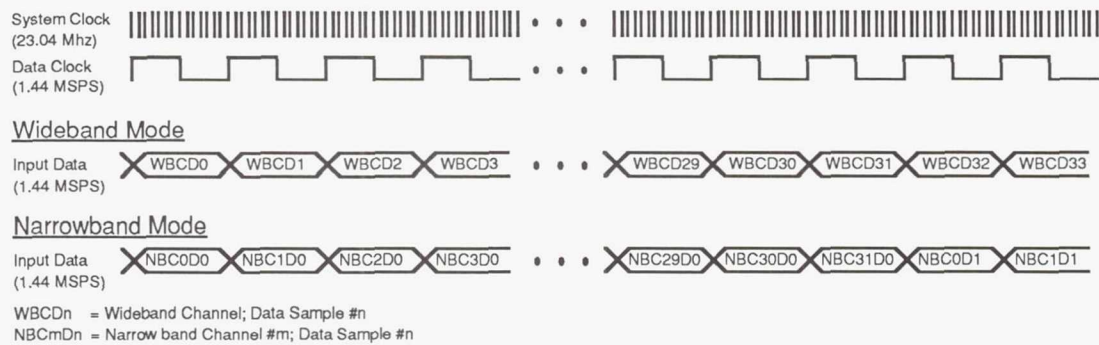
Figure 2. Input Sample/Data Clock/System Clock Timing

The MRD, shown conceptually in Figure 1, is comprised of five functional blocks: Re-Sample/Matched Filter, Derotate, Output, Symbol Sync, and Carrier Track.

The re-sampling filters interpolate the desired input waveform samples from the input channelizer samples. Interpolation is necessary to create samples at the desired sampling interval. The re-sampling filters also provide proper sample positioning, and compensation due to rate offsets. The interpolation filter is combined with a matched filter to accomodate pulse shaping of the baseband waveform.

Residual phase error and carrier frequency offsets are removed by derotation of the complex baseband waveform. The phase error estimate is generated in the carrier tracking loop.

Symbol synchronization is accomplished by means of a symbol sync loop to estimate the timing error, and provide the timing offset for interpolated samples. The current error estimator has been simplified for QPSK operation. The symbol sync also contains a NCO that perfoms the rate conversion for the interpolated samples.

Residual carrier phase is estimated by a second order, type-2, carrier tracking loop to provide a phase error that is then removed by derotation. Different modulation formats can be accommodated by simply modifying the phase error look-up table. The derotation process provides proper quadrature alignment.

Symbol decisions are made upon the midpoint samples generated through interpolation, with a serial data stream and corresponding data clock being output. The output data rate is 2.048 Mbit/sec.for the WBC's, and 64Kbit/sec. for the NBC's. Various modulation formats can be accommodated by modifying the symbol decision look-up table. The output data and clock will be from a single WBC or from 32 NBC's time-division-multiplexed onto the same data and clock line. Thirty-two separate data and clock lines for the NBC's are also available.

III.   Architecture

The MRD, a detailed block diagram shown in Figure 3, is comprised of the following processing elements:

- Input Buffer: To store input samples necessary for filter calculation.
- Re-Sampling Filters: For matched / interpolation filtering.
- Derotate: Complex Multiply needed for phase error removal.
- De-Stagger Buffer: To remove staggering and store samples.
- Symbol Timing Error Estimator: Estimates timing error of interpolated samples.
- Symbol Sync Loop Filter: Low-Pass Filters timing estimate.
- Symbol Sync NCO: Provides timing offset for interpolated samples.
- Phase Error Estimator: Estimates residual phase error.
- Carrier Tracking Loop Filter: Low-Pass Filters timing estimate.
- Carrier Tracking NCO: Tracks phase error to be removed by derotation.
- Symbol Decision: For symbol decision, and data/clock formatting.

274
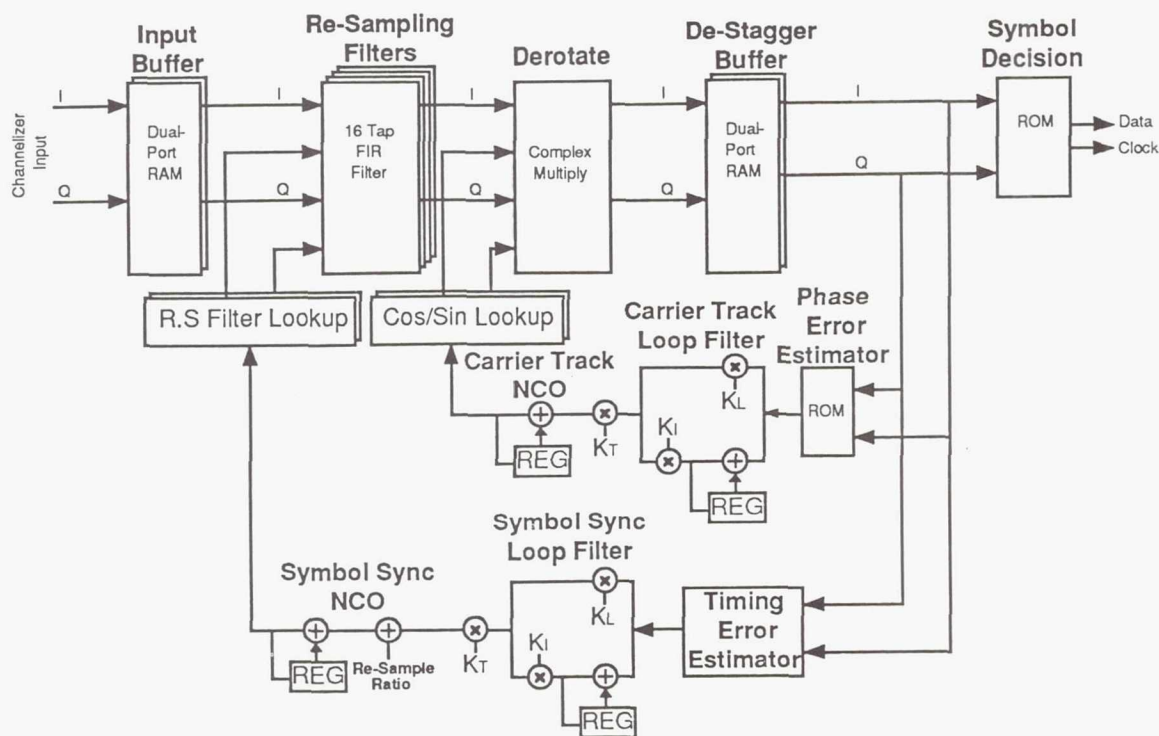
Figure 3. Multi-Rate Demodulator (Detailed)

1. Interpolation / Re-Sampling Filters

The input data is sampled at the Nyquist minimum sampling frequency, and therefore interpolation (or re-sampling) can be performed to generate any desired sample from the input baseband waveform (ref. 1). For the MRD a sample at the midpoint and at the transition point are needed for demodulator processing. The interpolation filter is combined with a matched filter to accommodate various baseband signal conditioning.

The MRD currently being developed has a WBC input data rate of 1.44 Msample/sec. and a NBC input rate of 45 Ksample/sec. This gives a 1.44 Mhz period (or a 16 cycle system clock input sample epoch) in which demodulator processing must be completed for a particular channel, if NBC, or input sample if WBC. The symbol rates, in absence of any offsets, are 1.024 Msymbol/sec. for the WBC and 32 Ksymbol/sec. for the NBC. For demodulator processing it is necessary to generate two samples per symbol giving an interpolated data rate of 2.048 Msample/sec. for the WBC and 64 Ksample/sec. for the NBC. The input waveform is "re-sampled" to give data samples at the desired sampling interval through interpolation of the input data points. The interpolation process is performed by a digital 16 tap (ref. 1) Finite-Impulse-Response (FIR) filter that is combined with a matched filter to accommodate pulse shaping of the baseband waveform. Pulse shaping currently implemented is a Square-Root-Raised-Cosine with filter coefficients being input from a look-up table. Different pulse shaping schemes can be implemented by modification of the filter coefficient look-up table. Two filter outputs are needed per quadrature arm (4 total) to generate the two interpolated output points during an input sample epoch.

All interpolation analysis for the WBC and NBC is identical since both the input and output sample rates for the WBC are 32 times that of the NBC's. Proper symbol synchronization, and position of interpolated samples is achieved through the Symbol Sync NCO that performs the rate conversion. The NCO will provide the 5 bit timing offset, representing ±1/2 an input sample in 32 steps, that the current sample being interpolated will have. The timing offset will select a set of filter coefficients to perform the interpolation. There are 32 sets of 16 filter coefficients, one

275

coeffiecient set for each posible timing offset, stored in a 512x8 look-up table. The interpolation filters can generate at most two samples, and at least one sample during an input sample epoch with the number of samples generated being determined by the NCO overflow status.

Since the MRD runs at a multiple of the input data rate of 45 Ksample/sec., with the interpolated data rate being 64 Ksample/sec. for NBC's (2.048 Msample/sec. for WBC's), samples out of the re-sampling filters will not be at a uniform rate, however the average rate over time will be 64 Ksamples/sec. This is illustrated in Figure 4.

mp = mid-point
t = transition point
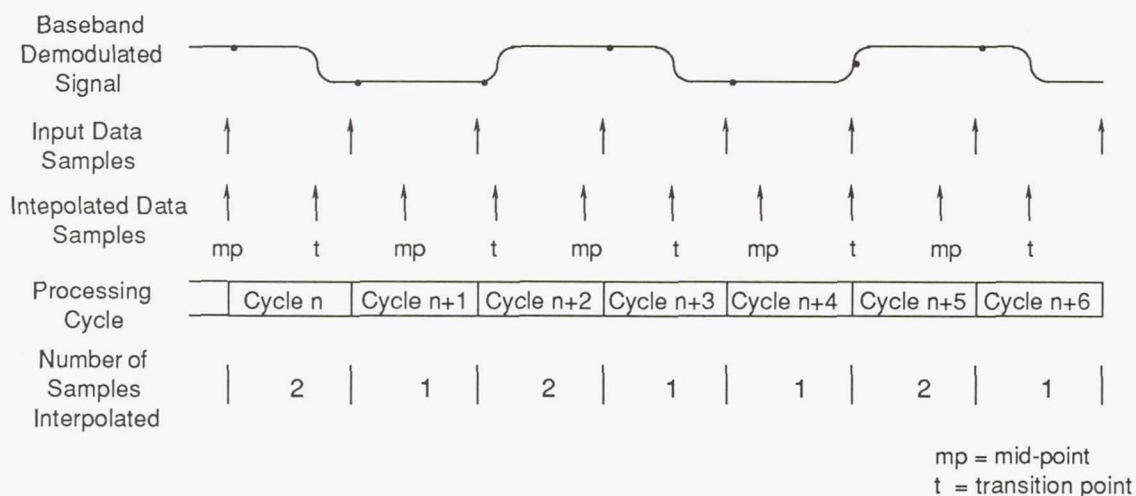
Figure 4. Re-Sampling Filter Interpolation Timing

2. De-Stagger

The modulation format, OQPSK, is such that I and Q samples are staggered by half a symbol period. Besides the staggering, the interpolation process will output data at an irregular rate. Since a previous midpoint, previous transition, and current midpoint are needed for demodulator processing, and due to the staggering and irregular rate, a memory buffer is necessary. If the necessary data samples are not resident in the de-stagger buffer, some processing for the current input sample epoch is bypassed. The symbol sync NCO and carrier tracking accumulator are still updated, but no symbol decisions or symbol timing/ phase error estimates will be made.

3. Symbol Synchronization Loop

Symbol syncronization loop is comprised of the timing error estimator, the Symbol Sync Loop Filter, and the Symbol Sync NCO.

The Timing Error Estimator generates a non-zero estimate if a transition has occurred between the previous and current midpoint samples. The estimates for both the I and the Q are summed and input to the Symbol Sync Loop Filter. Synchronization loops of this type are sometimes refered to as a Data-Transition-Tracking-Loop or DTTL (ref. 2).

In order to determine whether a transition has occurred, the previous and current midpoints are compared. If a transition did occur, the value of the transition sample, with its polarity adjusted as shown in Figure 5, will represent the error estimate. The end result is that if the transition point is early, a negative error estimate will be created to slow the NCO down, and if the transition point is late a positive estimate will be created to speed the NCO up.

The Symbol Sync Loop Filter acts as a low pass filter, so that depending on the loop bandwidth selected, RMS jitter due to thermal noise can be traded off for pull-in time and phase noise performance.

276

MSBn	MSBn-1	Error
0	0	0
0	1	+Tn-1
1	0	-Tn-1
1	1	0
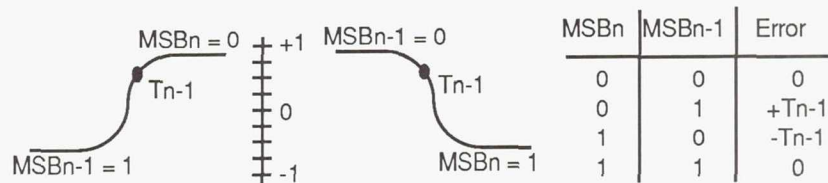
Figure 5. Timing Error Estimate

All loop filter coefficients (K_L, K_I and K_T) are powers of two to enable using shifts instead of multiplies and are programmable to accommodate different loop bandwidths. The word width of the loop filter output is 24 bits to allow for a wide range (from 10^{-1} • symbol rate to 10^{-5} • symbol rate) of possible loop bandwidths (ref. 1).

The output of the loop filter is input to the NCO to fine tune the NCO frequency. The coarse tune is determined by the re-sampling ratio. As mentioned previously, the NCO determines if one or two interpolated samples, for both I and Q, are generated during the current input sample epoch. If the NCO overflows on the first update, then the current input samples are only sufficient to generate one interpolated sample. If however the NCO does not overflow on the first update, it is updated again, and two interpolated samples will be generated. The most-significant 5 bits of the 24 bit NCO output are used as a re-sampling filter select, giving 32 posible timing offsets.

4. Carrier Tracking Loop

Carrier Tracking Loop (ref. 3) is comprised of the Phase Error Estimator, the Carrier Tracking Loop Filter, and a Carrier Tracking NCO (or accumulator).

The Carrier Tracking Estimator uses the current midpoint from both the I and Q arm to estimate the phase error. This estimate is computed as:

$$\theta_{error}(I,Q) = TAN^{-1}(Q/I) - \theta_{optimum} \tag{1}$$

where $\theta_{optimum}$ is computed depending on which quadrant the current symbol resides as shown in Figure 6.
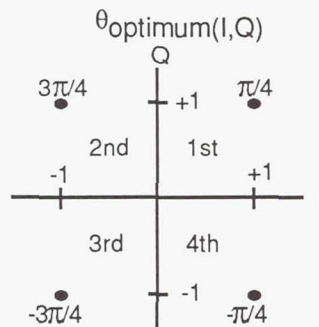
Figure 6.

The phase error estimate is computed by means of a look-up table (1024 x 8) that could be modified to accommodate different modulation schemes.

The phase error estimate is input to the Carrier Tracking Loop Filter which is of the same type as the Symbol Sync Loop Filter. The output of the loop filter is input to the 24 bit Carrier Tracking NCO which will track the phase error of the input waveform. The most-significant 8 bits of the NCO output will be mapped into $\pm\pi$ radians and will be used as an address to a sin/cos look-up

table (256 x 8). A complex multiply is necessary to remove the estimated phase error from the baseband waveform as shown below:

$$X_{de\text{-}rot} = X_{rot} \cdot e^{-j\theta est} \qquad (2)$$

$$= [X_{rotI} \cdot \cos\theta_{est} + X_{rotQ} \cdot \sin\theta_{est}] + j[X_{rotQ} \cdot \cos\theta_{est} - X_{rotI} \cdot \sin\theta_{est}] \qquad (3)$$

where,

$X_{de\text{-}rot}$ = Derotated complex baseband waveform
X_{rot} = Rotated input complex baseband waveform
X_{rotI} = Real componant of X_{rot}
X_{rotQ} = Imaginary component of X_{rot}
θ_{est} = phase error estimate from carrier tracking accumulator

5. Symbol Decision

The output symbol decisions are made upon the I and Q arm independently using the previous and current midpoint samples. The data is differentially encoded with the decision being made on the most-significant (sign) bit of the two midpoints. OQPSK will generated two bits per symbol, one for the I arm and one for the Q arm. The data bits are muxed onto the same line to create a serial data stream, with a corresponding data strobe. The data for the NBC's can be separated into 32 independent data and clock lines, or muxed onto the same data and clock lines in a TDM'ed fashion. A corresponding index is also output for the TDM'ed clock and data to determine the current channel. The symbol decision look-up can be modified to accommodate various modulation formats.

IV. Conclusions

The digital Multi-Rate Demodulator with many modulation dependent functions being performed in look-up tables, has a high degree of inherent flexibility. This flexibility will allow the consideration of different modulation formats without hardware modification. The MRD currently being developed is a prototype for definition of a future Application-Specific-Integrated-Circuit (ASIC) implementation. The ASIC implementation will be a compact, low power unit for insertion into future satellite systems with overall system channel capacity easily expanded by the addition of MRD modules.

The digital Multi-Rate Demodulator architecture, with it's high degree of flexibility can accommodate a wide variety of modulation formats and symbol rates without design modification. The MRD provides an efficient, low-complexity, digital implementation that can be suited to a host of different satellite applications.

References

1. R.E. Crochiere and L.R. Rabiner, *Multirate Digital Signal Processing.* Englewood Cliffs, N.J.: Prentice-Hall, 1983, pp. 13-58.

2. W.L. Lindsey and M.K. Simon, *Telecommunications Engineering.* Englewood Cliffs, N.J.: Prentice-Hall, 1973, pp. 442-457.

3. J.K. Holmes, *Coherent Spread Spectrum Systems.* New York, N.Y.: John Wiley and Sons, 1982, pp. 153-120.

MULTI-STAGE DECODING

OF

MULTI-LEVEL MODULATION CODES[1]

Shu Lin
Department of Electrical Engineering
University of Hawaii at Manoa
Honolulu, Hawaii 96822, U.S.A.

Tadao Kasami
Faculty of Engineering Science
Osaka University
Toyonaka, Osaka 560, Japan

Daniel J. Costello, Jr.
Dept. of Electrical and Computer Engg.
University of Notre Dame
Notre Dame, IN 46556

Abstract

This paper investigates various types of multi-stage decoding for multi-level modulation codes. It is shown that if the componenet codes of a multi-level modulation code and types of decoding at various stages are chosen properly, high spectral efficiency and large coding gain can be achieved with reduced decoding complexity. Particularly, it is shown that the difference in performance between the suboptimum multi-stage soft-decision maximum likelihood decoding of a modulation code and the single-stage optimum soft-decision decoding of the code is very small, only a fraction of dB loss in SNR at BER of 10^{-6}.

1. Introduction

Coded modulation is a technique of combining coding and bandwidth efficient modulation to produce modulation (or signal space) codes for achieving reliable data transmission without compromising bandwidth efficiency [1-4]. Over the last eight years, a great deal of research effort has been expended in constructing good bandwidth efficient modulation codes. Among all the proposed methods for constructing modulation codes, the most powerful one is the multi-level construction method [2,3,5-9]. This method allows us to construct modu-

---

lation codes systematically with arbitrarily large minimum **squared Euclidean distance** from Hamming distance component codes (**binary** or **nonbinary**, **block** or **convolutional**) in conjunction with proper **bits-to-signal** mapping through **signal set partitioning**. If the component codes are chosen properly, the resultant multi-level modulation code not only has good minimum squared Euclidean distance but is also rich in structural properties such as: **regularity, linear structure, phase symmetry** and **trellis structure**. These structural properties simplify the error performance analysis, encoding and decoding implementations, and resolution of carrier-phase ambiguity. A major advantage of multi-level modulation codes is that these codes can be decoded in **multiple stages** with component codes decoded sequentially stage by stage, with decoded information passed from one stage to another stage. Since component codes are decoded one at a time, it is possible to take advantage of the structure of each component code to simplify the decoding complexity and reduce the number of computations at each stage. As a result, the overall complexity and number of computations needed for decoding a multi-level modulation code will be greatly reduced. This allows us to achieve high reliability, large coding gain and high spectral efficiency with reduced decoding complexity.

2. Multi-Stage Decoding of Multi-Level Modulation Codes

There are four possible types of multi-stage decoding:

(1) **Multi-stage Soft-decision Maximum Likelihood Decoding** - Each stage of decoding is a soft-decision maximum likelihood decoding;

(2) **Multi-stage Hard-decision Maximum Likelihood Decoding** - Each stage of decoding is a hard-decision maximum likelihood decoding;

(3) **Multi-stage Bounded-distance Decoding** - Each decoding stage is a bounded-distance decoding based on a certain distance measure, e.g., Hamming distance; and

(4) **Hybrid Multi-stage Decoding** - Mixed types of decoding are used among the stages.

With the multi-stage soft-decision maximum likelihood decoding, each component code of a multi-level modulation code is chosen to have trellis structure and is decoded with the soft-decision Viterbi decoding algorithm. Since the decoding at each stage depends on the decoded information from the previous decoding stages, there is a likelihood of error propagation. As a result, the overall decoding is not optimum even though the decoding at each stage is optimum. It is a suboptimum decoding. However, the error propagation effect can be made negligibly small, if the first few component codes(mostly the first component code) of a multi-level modulation code are powerful. Based on our analysis and simulation of the error performance of several efficient multi-level modulation codes, we find that the difference in performance between the suboptimum multi-stage decoding and the single-stage

optimum decoding is very small, only a fraction of dB loss in SNR at the BER(block or bit error rate) of 10^{-6}.

With the multi-stage hard-decision maximum likelihood decoding, each component code is also chosen to have trellis structure, but is decoded with the hard decision Viterbi decoding algorithm. This type of multi-stage decoding further simplifies the decoding complexity, however there is a 2-2.5 dB **loss** in SNR compared to the optimum soft-decision decoding. Even with some loss in SNR, the multi-stage hard-decision maximum likelihood decoding still achieves significant coding gain over an uncoded system with the same spectral efficiency based on our computations and simulations of error performance of some multi-level modulation codes.

With the multi-stage bounded distance decoding, component codes of a multi-level modulation code are decoded with bounded-distance decoding based on either Euclidean or Hamming distance measure. If a component code is binary, its minimum squared Euclidean distance is linearly propotional to its minimum Hamming distance. As a result, it can be decoded based on its minimum Hamming distance. In this case, algebraic or majority-logic decoding may be used. Results show that if the first-level component code is a low-rate powerful code and the other component codes are high-rate code, the multi-stage bounded distance decoding can also achieve significant coding gain over an uncoded system without any bandwidth expansion and with greatly reduced decoding complexity.

The hybrid multi-stage decoding provides an excellent trade-off between coding gain and decoding complexity. With this scheme, the lower-level decoding stages (specially the first-level decoding) are soft-decision maximum likelihood decoding using Viterbi decoding algorithm and the higher-level decoding stages are hard-decision maximum likelihood or bounded distance decoding. Based on our computation and simulation of error performance of some multi-level modulation codes, we find that the hybrid multi-stage decoding has less than one dB loss in coding gain compared to the optimum decoding.

A very natural architecture for a multi-stage decoder is the **pipeline architecture**. For a multi-level modulation code with m component codes, the decoder is organized to decode m received vectors in pipeline process. While the decoder is decoding the m-th component vector of the earliest received vector in the pipe, it is also decoding the $(m-1)$-th component vector of the next received vector in the pipe, ..., and the first component vector of the most recent received vector. This pipeline architecture speeds up the decoding process.

3. Examples

Consider a basic 3-level 8-PSK block modulation code of length 32 with the following three component codes: (1) C_1 is the $(32,6)$ Reed-Muller code with Hamming distance $\delta_1 = 16$; (2) C_2 is the $(32,26)$ Reed-Muller code with Hamming distance $\delta_2 = 4$; and (3) C_3 is the $(32,31)$ even parity check code with Hamming distance $\delta_3 = 2$. This basic 3-level 8-

PSK modulation code, $C = C_1 * C_2 * C_3$, has minimum squared Euclidean distance $D[C] = 8$ and spectral efficiency $\eta[C] = 63/32 = 1.966$. This code achieves 6 dB asymptotic coding gain over the uncoded QPSK with optimal decoding. The first component code C_1 has a 4-section 16-state trellis, the second component code C_2 also has a 4-section 16-state trellis, and the third component code C_3 has a 32-section 2-state trellis. The overall modulation code $C = C_1 * C_2 * C_3$ has a 512-state trellis. To perform the single-stage optimum decoding for the overall code, we need to build a soft-decision Viterbi decoder with 512 states which is quite complex and expensive. However, with the multi-stage soft-decision maximum likelihood decoding for this code, we need only two 16-state and one 2-state Viterbi decoders (a total of 34 states) for the three component codes. The total complexity is much less than that of a single 512-state Viterbi decoder for optimum decoding. The error performance of the code is shown in **Figure 1.** We see that, with multi-stage soft-decision maximum likelihood decoding, there is almost 5 dB in **real coding gain** over the uncoded QPSK at block-error-rate (BER) 10^{-6}, which is only 1 dB away from the 6 dB asymptotic coding gain. If optimum decoding is performed, the real coding gain of the code over the uncoded QPSK is 5.25 dB at BER $= 10^{-6}$. We see that there is an **excellent trade-off** between the error performance and decoder complexity.

Figure 1 also includes the error performance of the above 3-level 8-PSK modulation code using 3-stage hard-decision maximum likelihood decoding. We see there is a 2.3 dB loss in SNR at the BER of 10^{-6} compared with the 3-stage soft-decision suboptimum decoding. However, there is still 2.7 dB coding gain over the uncoded QPSK system with very little bandwidth expansion. With the 3-stage hard-decision decoding, the decoding complexity is further reduced.

As a second example, consider a 3-level 8-PSK block modulation code of length 64 with the following component codes: (1) C_1 is the second order (64,22) Reed-Muller code with minimum Hamming distance $\delta_1 = 16$; (2) C_2 is the 4-th order (64,57) Reed-Muller code with minimum Hamming distance $\delta_2 = 4$; and (3) C_3 is the (64,63) even parity check code with minimum Hamming distance $\delta_3 = 2$. This 3-level 8-PSK modulation code, $C = C_1 * C_2 * C_3$, has minimum squared Euclidean distance $D[C] = 8$ and spectral efficiency $\eta[C] = 142/64 = 2.22$. The first component code has a 4-section trellis diagram with 2^{10} states, the second component code has a 4-section trellis diagram with 2^5 states, and the third component code has a 2-state trellis diagram. The overall code has a 4-section trellis diagram with 2^{16} states. Decoding this code with the single-stage soft-decision maximum likelihood decoding using Viterbi algorithm is prohibitively complex. However, with 3-stage soft-decision maximum likelihood decoding, this code achieves a 4.5 dB coding gain over the uncoded QPSK system at the block-error-rate 10^{-6} (see Figure 2) with a big reduction in decoding complexity(from a complexity of 65536 states to a complexity of 1058 states). In fact, this coding gain is achieved with a bandwidth reduction. With the 3-stage hard-decision bounded distance decoding, the code also achieves significant coding gain over the uncoded QPSK system with bandwidth reduction(see Figure 2). There is a 2.2 dB loss in coding gain compared with

the 3-stage soft-decision maximum likelihood decoding, however the decoding complexity is greatly reduced. Note that the first component code is majority-logic decodable and the second component code is simply a distance-4 extended Hamming code which can be easily decoded. To improve the performance while still keeping the complexity down, we may use the hybrid multi-stage decoding in which the first component code is decoded with the hard-decision bounded distance decoding, and the second and third componenet codes are decoded with the soft-decision maximum likelihood decoding using the Viterbi algorithm.

4. Conclusion

In our examples, we used block modulation codes to demonstrate the effectiveness of the multi-stage decoding. The multi-stage decoding can be applied to decode the multi-level trellis modulation codes. This type of decoding for multi-level modulation code really offers the best of three worlds, spectral efficiency, coding gain(or error performance), and decoding complexity.

REFERENCES

1. Ungerboeck, G., "Channel Coding with Multilevel/Phase Signals," **IEEE Trans. on Information Theory**, Vol. IT-28, No. 1, pp. 55-67, January 1982.

2. Imai, H. and Hirakawa, H., "A New Multilevel Coding Method Using Error Correcting Codes," **IEEE Trans. on Information Theory**, Vol. IT-23, No. 3, pp. 371-376, May 1977.

3. Ginzburg, V.V., "Multidimensional Signals for a Continuous Channel," **Problemy Peredachi Informatsii**, Vol. 20, No. 1, pp. 28-46, 1984.

4. Forney, G.D., Gallager, R.G., Lang, G.R., Longstaff, F.M., and Quereshi, S.U., "Efficient Modulation for Band-limited Channels," **IEEE J. Selected Areas in Communications**, Vol. SAC-2, pp. 632-646, September 1984.

5. Sayegh, S.I., "A Class of Optimum Block Codes in Signal Space," **IEEE Trans. on Communications**, Vol. COM-30, No. 10, pp. 1043-1045, October 1986.

6. Tanner, R.M., "Algebraic Construction of Large Euclidean Distance Combined Coding Modulation Systems," **Abstract of Papers, 1986 IEEE International Symposium on Information Theory**, Ann Harbor, October 6-9, 1986, also **IEEE Trans. on Information Theory**, to be published.

7. Pottie, G.J. and Taylor, D.P., "Multi-Level Channel Codes Based on Partitioning," **IEEE Trans. on Information Theory**, Vol. IT-35, No. 1, pp. 87-98, January, 1989.

8. A.R. Calderbank, "Multi-Level Codes and Multi-Stage Decoding," **IEEE Trans. on Communications**, Vol. COM-37, No. 3, pp. 222-229, March 1989.

9. Kasami, T., Takata, T., Fujiwara, T., and Lin, S., "On Multi-Level Block Modulation Codes," **IEEE Trans. on Information Theory**, Vol. IT-37, No. 4, July 1991.

Figure 1

Figure 2

285
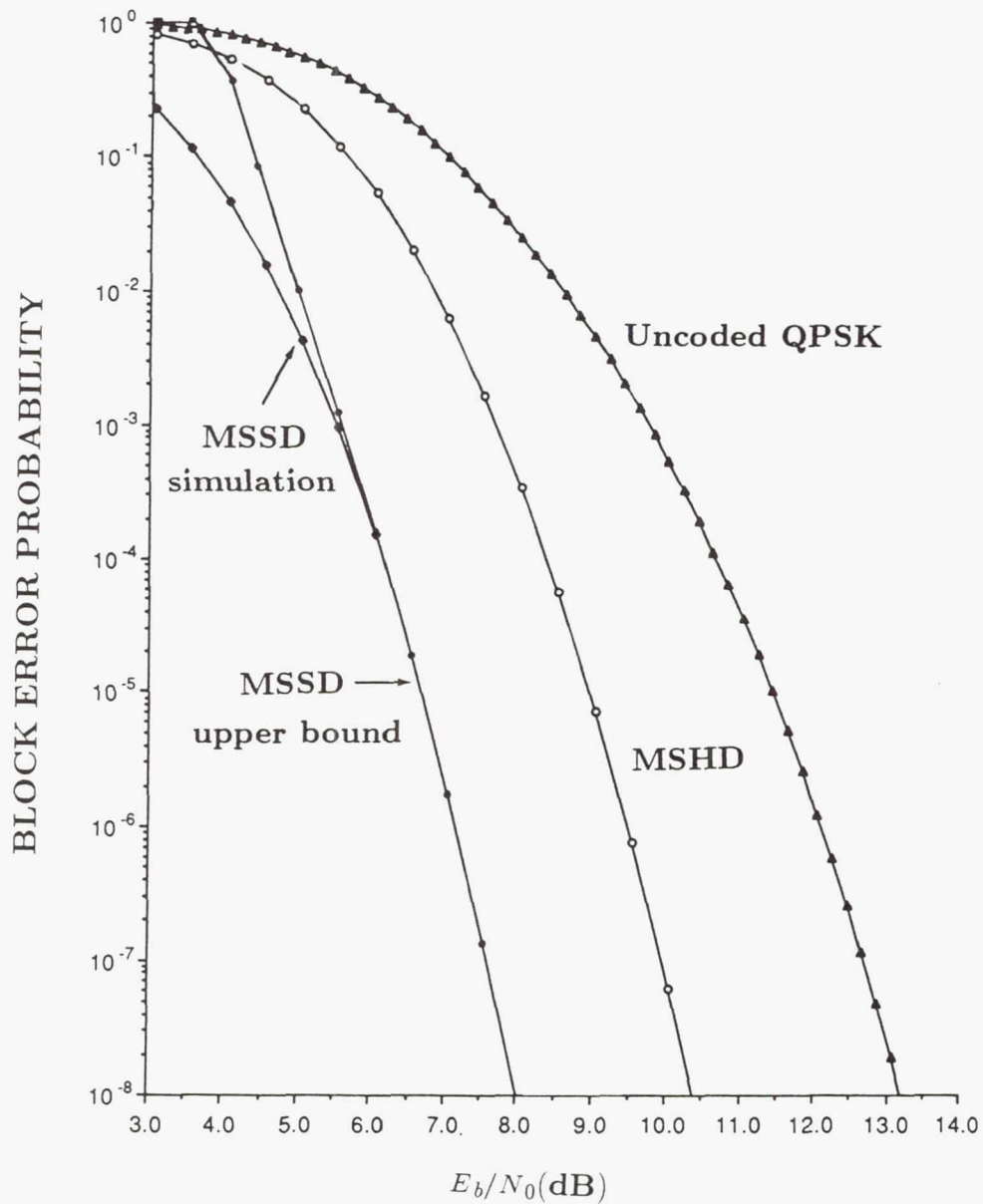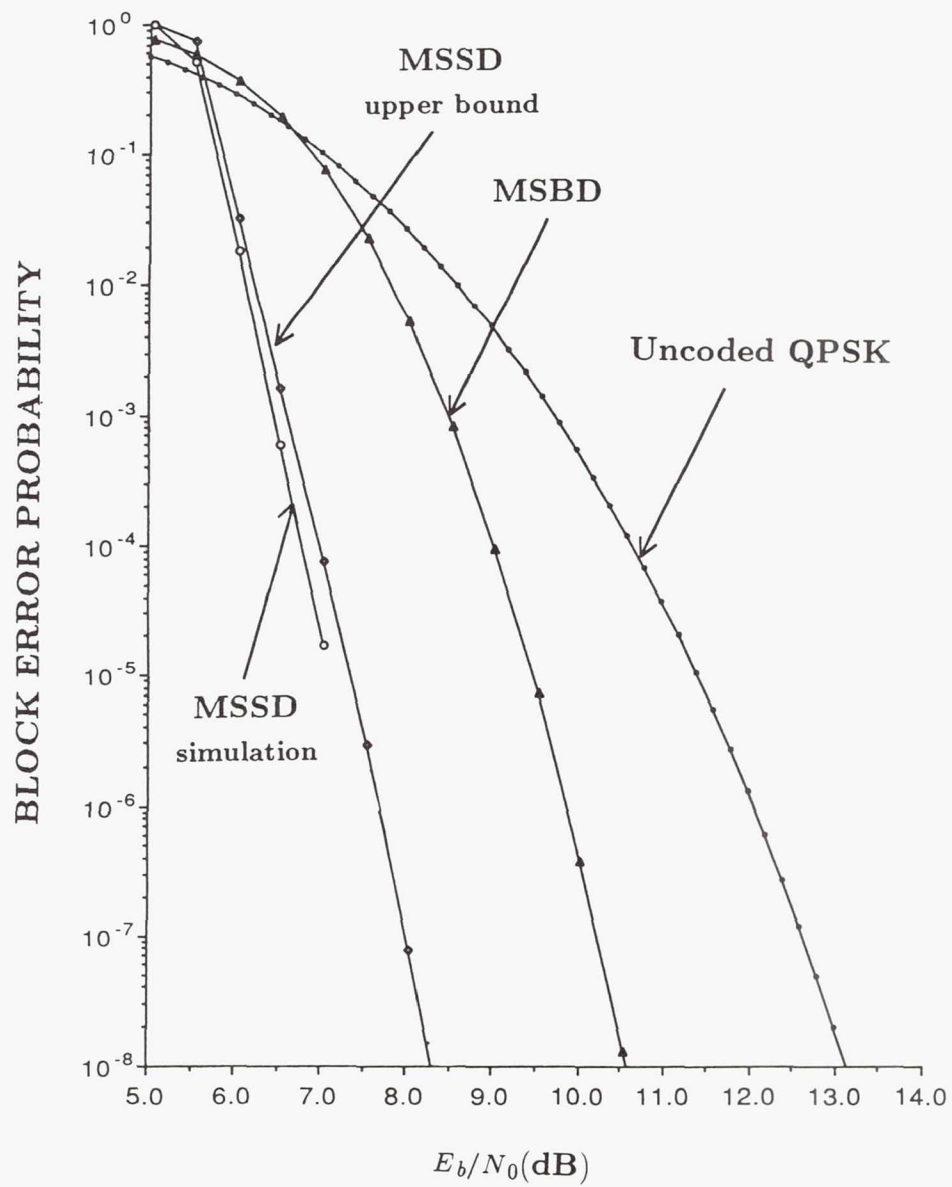
Page intentionally left blank

BASEBAND PULSE SHAPING TECHNIQUES FOR NONLINEARLY AMPLIFIED π/4-QPSK AND QAM SYSTEMS

KAMILO FEHER
Digital Communications Research Laboratory,
Department of Electrical and Computer Engineering
UC DAVIS
University of California, Davis
Davis, CA 95616

ABSTRACT

A new generation of multi-state π/4-shifted QPSK and of Superposed Quadrature-Amplitude-Modulated (SQAM) modulators-coherent demodulators (modems) and of Continuous Phase Modulated (CPM)-Gaussian Premodulation Filtered Minimum-Shift-Keying (MGMSK) systems is proposed and studied. These modems will lead to bandwidth and power efficient satellite communications systems designs. As an illustrative application, a new baseband processing technique π/4-Controlled Transition PSK (π/4-CTPSK) is described. To develop a cost and power efficient design strategy, we assume that nonlinear, fully saturated high power amplifiers (HPA) are utilized in the satellite earth station transmitter and in the satellite transponder. Modem structures which could lead to Application-Specific-Integrated-Circuit (ASIC) satellite on-board processing "universal" modem applications are also considered.

Multistate GMSK (i.e., MGMSK) signal generation methods by means of two or more RF combined nonlinearly amplified SQAM modems and by one multistate (in-phase and quadrature-baseband premodulation filtered-superposed) SQAM architecture and one RF nonlinear amplifier are studied. During the SQAM modem development phase we investigate the potential system advantages of the π/4-shifted logic (such as used in the U.S. digital cellular standard π/4-DQPSK). The bandwidth efficiency of the proposed multistate GMSK and baseband filtered PAM-FM modulator (a new class in the CPM family) will be significantly higher than that of conventional G-MSK systems. To optimize the practical $P_e = f(E_b/N_0)$ performance we consider improved coherent demodulation MGMSK structures such as "deviated-frequency locking" coherent demodulators.

For relative low bit rate SATCOM applications, e.g., bit rates less than 300 kb/s, phase noise tracking-cancellation (for fixed site earth station) and phase noise cancellation as well as Doppler compensation (for satellite to mobile earth station) applications may be required. We study "digital channel sounding" methods which could cancel the phase noise-caused degradations of CPM and GMSK modems. The spectral-bandwidth efficiency of the proposed new class of modems will be in the 2b/s/Hz to 5b/s/Hz range with an anticipated out-of-band adjacent channel interference (ACI) in the ACI > -30dB to -40dB range. Hardware design and experimental optimization of a coherent multistate GMSK-SQAM structure will be initiated during the 1991-92 academic year. In this paper some of our preliminary research results, as of August 1991, are highlighted.

1. π/4-QPSK AND GMSK MODEMS

Modems suitable for digital satellite communications systems and for land mobile applications should satisfy at least the following requirements:

- A power efficient nonlinear amplifier may be used without introducing significant ACI, thus enabling efficient spectral utilization.

- A detection scheme able to achieve fast synchronization (including coherent and differential/discriminator detection) and good BER performance should be used.

In the search for modulation schemes which satisfy these requirements, numerous authors studied the "π/4-shifted QPSK systems," see Figure 1 through Figure 4. In this scheme, envelope fluctuations are significantly reduced compared to QPSK [Ref. 1]. Also, noncoherent detection schemes may be successfully applied to π/4-QPSK modulated signals. It has been shown that this modulation scheme achieves **double the spectral efficiency** of comparable constant envelope modulation schemes such as Gaussian Filtered MSK or GMSK. Note that GMSK has been adopted as the standard modulation format for the European DECT and GSM personal communications/cellular systems [Ref. 1; 4]. Due to the attractive features of π/4-QPSK, it has been chosen as the standard modulation scheme for the planned U.S. digital mobile radio system.

2. NEW IMPROVED MODEMS - DEVELOPED AT UC DAVIS

Experimental measurements at UC Davis and numerous other research laboratories indicate that in fully saturated, nonlinearly amplified systems π/4-QPSK has a significant spectral restoration, see Figure 5 [Ref. 1; 2; 3; 5]. For this reason we initiated a research program with an objective to maintain the attractive properties of π/4-QPSK and also to reduce the spectral spreading. In Figure 6 a new reduced spectrum π/4-QPSK modem, based on the SQAM concept is illustrated [Ref. 2]. In Figure 7 a particular baseband processed waveshape for CTPSK (controlled transition PSK) is illustrated [Ref. 3].

In an effort to reduce the envelope fluctuation of π/4-QPSK, a sinusoidal shaping scheme was introduced by Katoh-Feher in [Ref. 2]. This scheme sinusoidally shapes the phase transitions to reduce envelope fluctuations. The spectral advantages are shown in Figure 6. In our CTPSK scheme, in addition to a shaping technique, we also offset data transitions in the I and Q channels to further smooth the resulting phase transitions [Ref. 3], see Figure 7. The NLA (nonlinearly amplified) spectral advantages of our π/4-CTPSK, as compared to conventional π/4-QPSK are shown in Figure 8. An intuitive indication for this spectral reduction, after NLA, is the reduced envelope constellation diagram of the π/4-CTPSK signal, shown in Figure 9.

In Figure 10, we illustrate the NLA concept extended for SQAM and π/4-CTPSK systems to 64-QAM configurations. In our previous research we demonstrated that for conventional SQAM the BER = $f(E_b/N_0)$ results are close to theoretical performance, even in saturated NLA systems [Ref. 4], see Figure 13 and [Ref. 4 and 7]. For phase noise and Doppler shift compensation of relatively low bit rate mobile systems, we initiated the study of digital and analog pilot aided phase noise compensated modems [Ref. 1], see Figure 12.

For ASIC (Application-Specific-Integrated-Circuit) satellite on-board processing "universal" modem applications we are also studying the CPM-GMSK type of structures, illustrated in Figure 11. The resemblance between the GMSK and the QAM (SQAM) and QPSK transmitters/receivers could lead to universal ASIC implementations.
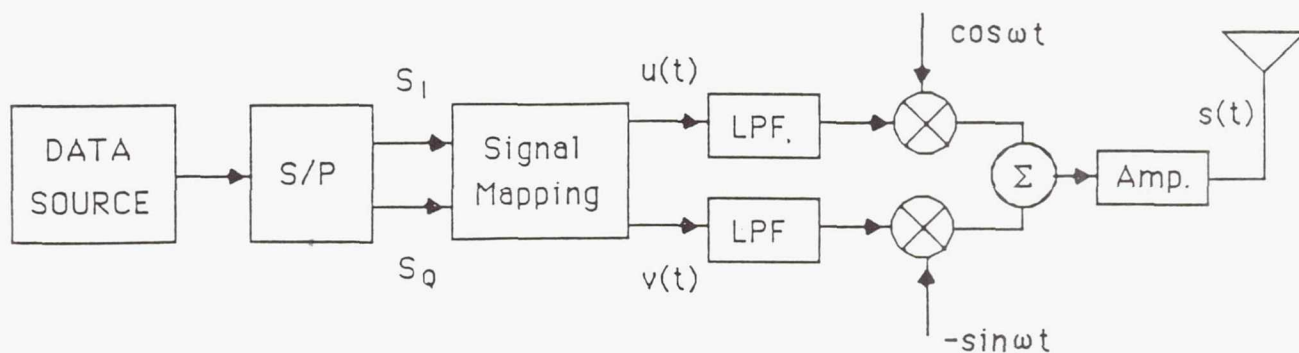
Figure 1. Block diagram of the transmitter of the π/4-QPSK modem.
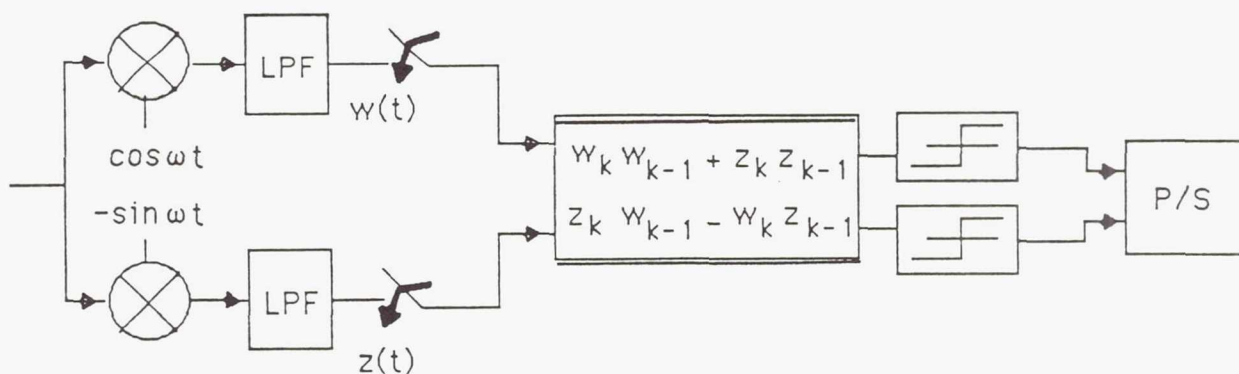
Figure 2. Block diagram of a π/4-QPSK coherent demodulator.
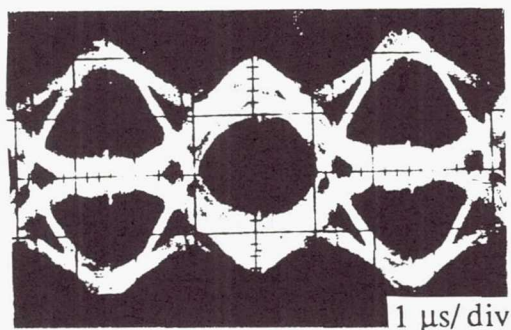
Figure 3. The "five-level" eye-diagram of coherent demodulated π/4-QPSK signals. At every other sampling instant the signals are two level. In between, the signals are three-level. Only the in-phase channel is shown. The bit rate used in this experimental setup is 800 kb/s.
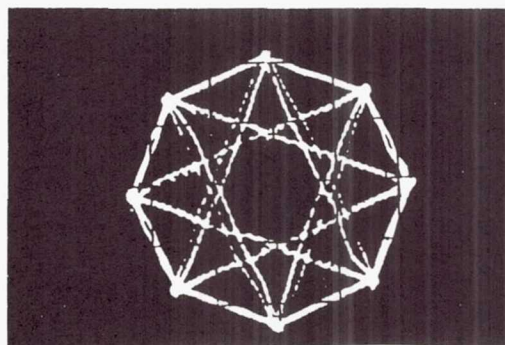
Figure 4. Constellation of the π/4-QPSK signal. In this hardware experiment, sine wave shaping π/4-QPSK (SP-QPSK) is used.
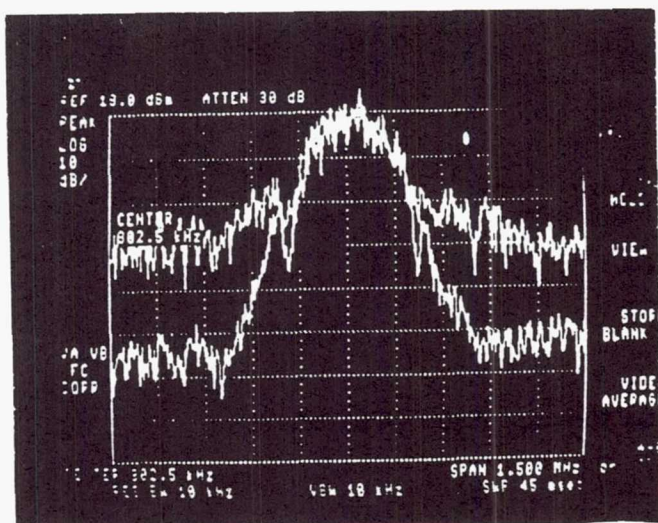
Figure 5. Experimental measurement results of the power spectral density of non-linearly amplified band limited π/4-QPSK. f_b = 400 kb/s, f_c = 1.18 MHz, Horizontal scale: 200 kHz/div, Vertical scale: 10 dB/div.

(a) Upper trace: saturated amplifier.

(b) Lower trace: amplifier input back-off 2dB measured on a prototype modem designed at UC Davis.
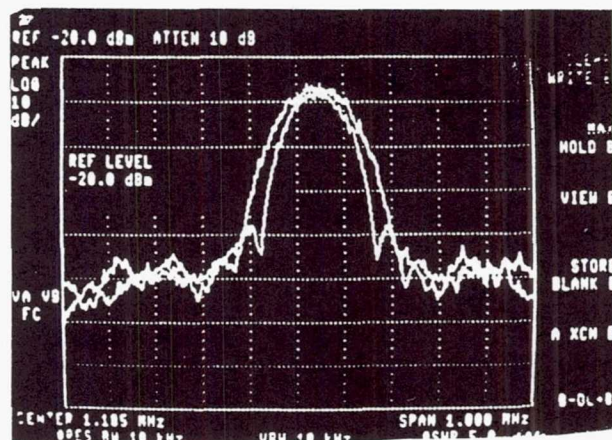
Figure 6. A new generation of improved performance π/4-QPSK modems developed at UC Davis has a reduced out-of-band spectrum in nonlinearly amplified (power efficient) radio system applications. In the illustrated spectral measurement we use an f_b = 250 kb/s rate. These "π/4-QPSK" (upper trace) and "SQAM" modems (lower trace) are described in [2; 14; 19; 21; and 24]. Horizontal scale: 100 kHz/div; vertical: 10 dB/div.
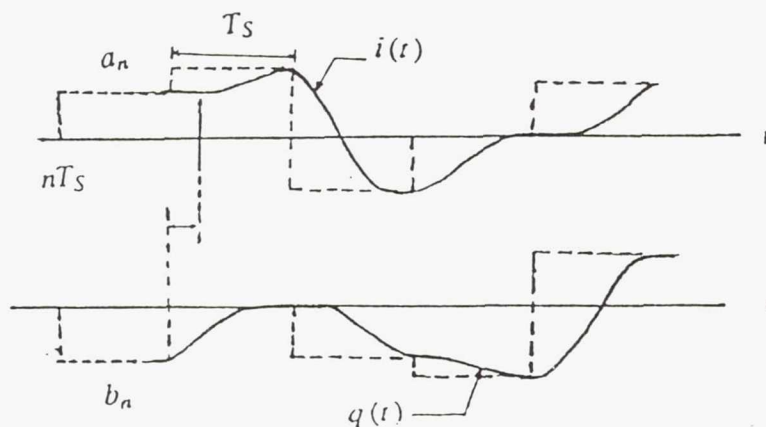
Figure 7. An illustration of π/4 - CTPSK processing.
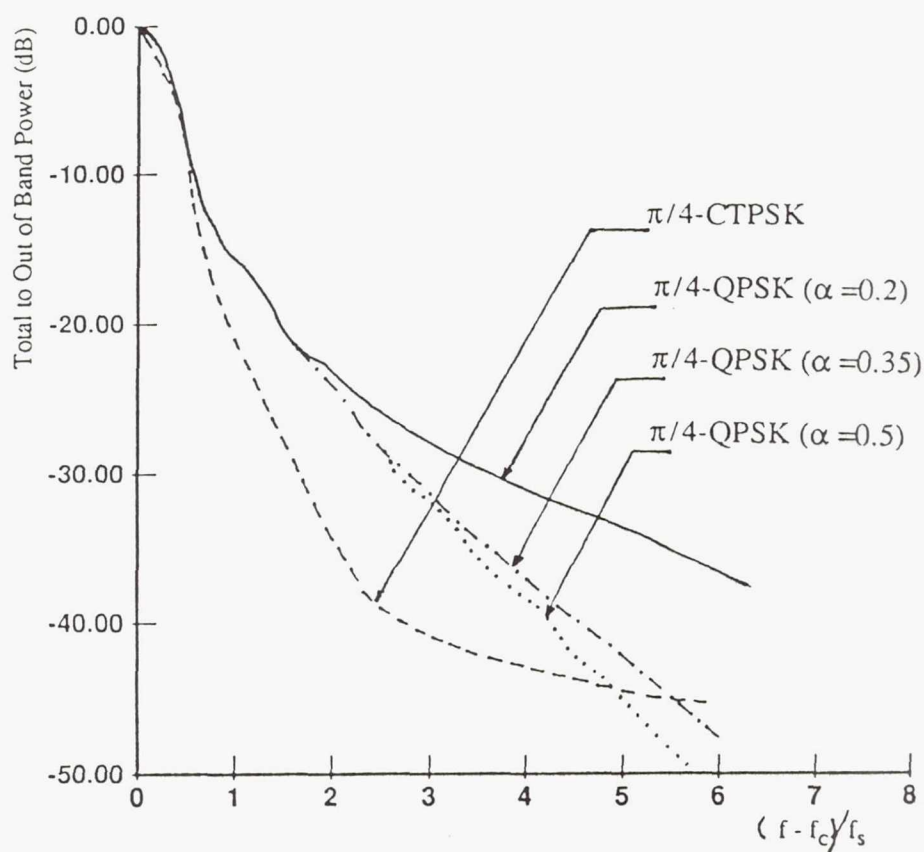
Figure 8. Comparison of total to out of band power ratios of $\pi/4$-CTPSK and $\pi/4$-QPSK with $\alpha = 0.2, 0.35$ and 0.5 (Hard-limited channel).
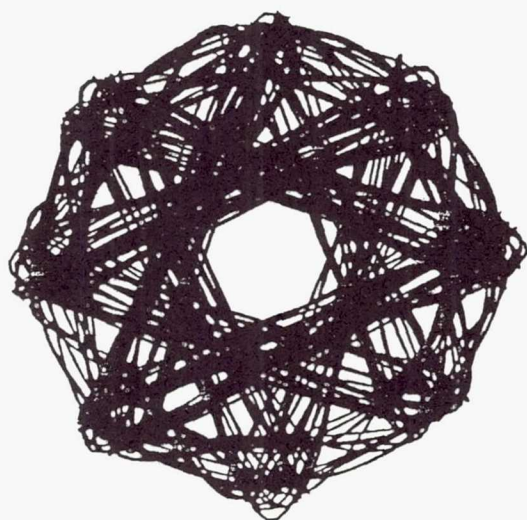
(c) $\pi/4$-QPSK ($\alpha = 0.5$)
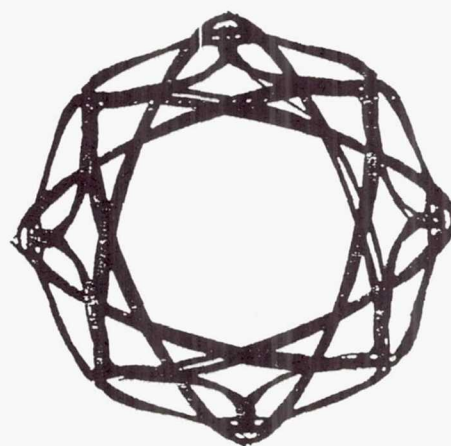
(d) $\pi/4$-CTPSK

Figure 9. Signal space diagrams (signal trajectories) of $\pi/4$-QPSK and $\pi/4$-CTPSK.

291
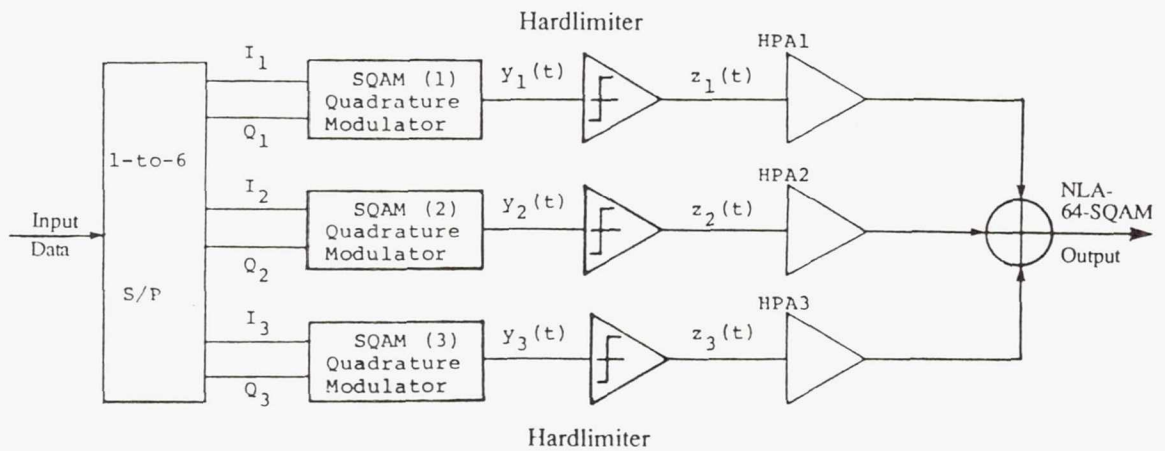
Figure 10. Block diagram of the NLA-64SQAM modulator. HPA1, HPA2 and HPA3 are operated in saturation mode.
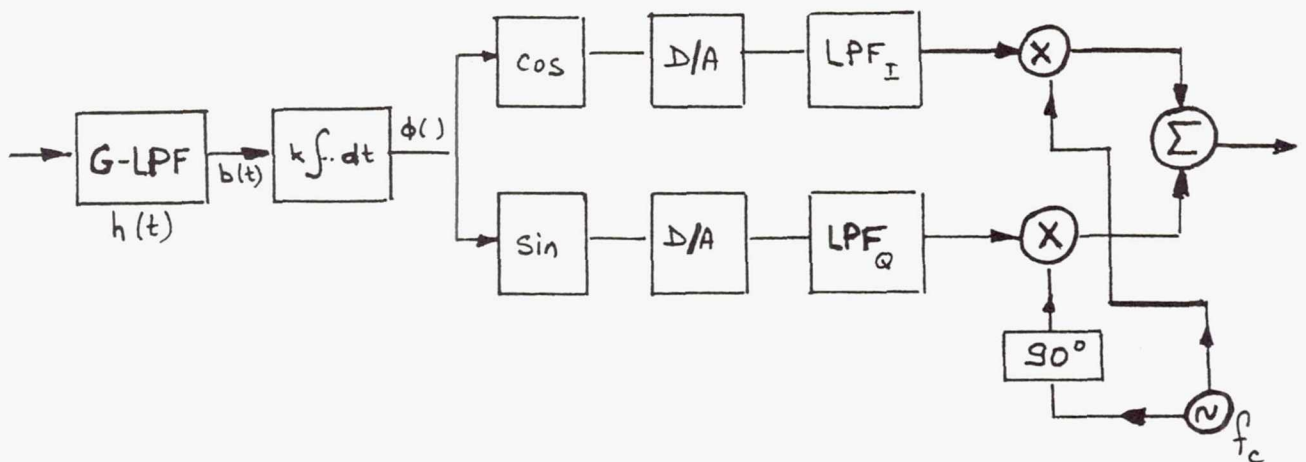
Figure 11. Continuous Phase Modulated (CPM) - Multistate Gaussian MSK
G-MSK Block Diagram

292
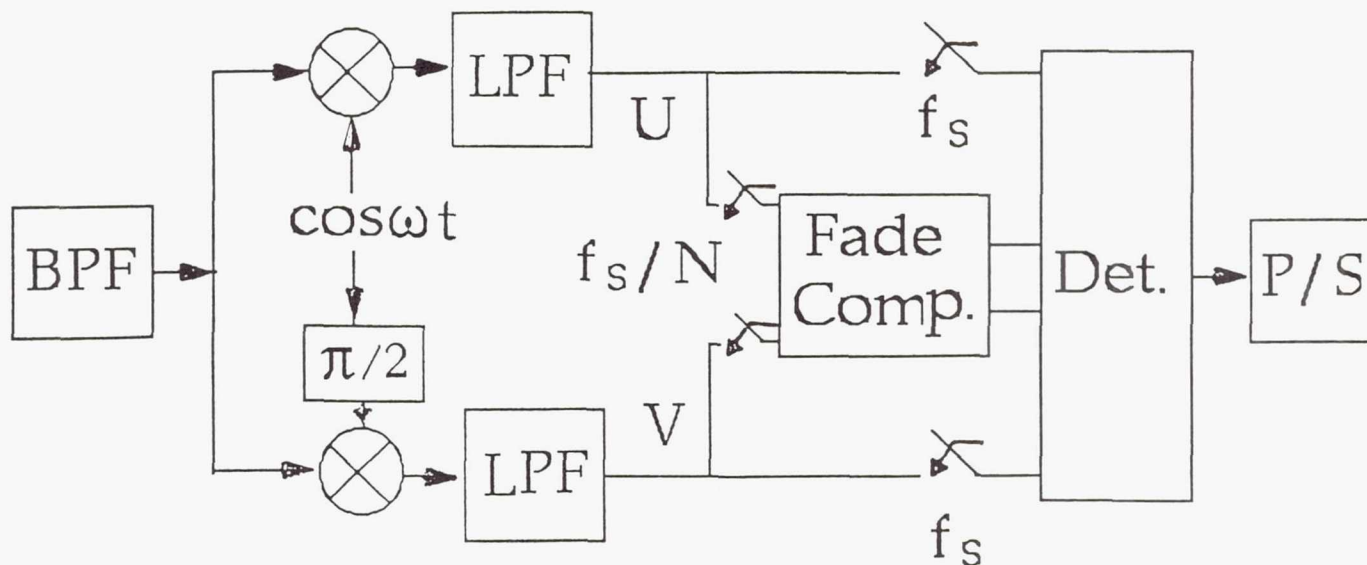
Figure 12. block diagram of the receiver of phase noise and Doppler shift PSK systems. The fade compensation block estimates the random phase caused by the Doppler spread. This random phase is subtracted from the detected phase of the signals in the decision block.
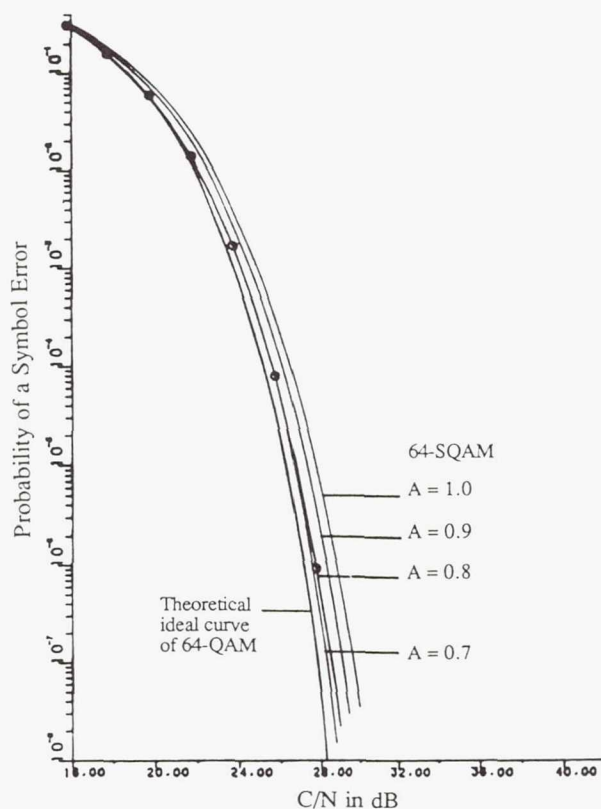
Figure 13. P(e) performance of a 64SQAM modem in an AWGN linear channel. Note: 4th-order Butterworth LPFs are used in the receiver [Ref. 4 and 7].

293

REFERENCES

1.   K. Feher: "Modems for Emerging Digital Cellular Mobile Radio Systems," (invited paper), *IEEE Transactions on Vehicular Technology*, May 1991.

2.   H. Katoh, K. Feher: "SP-QPSK: A New Modulation Technique for Satellite and Land-mobile Digital Broadcasting," *IEEE Transactions on Broadcasting*, Vol. 36, September 1990, pp. 195-202.

3.   D. Subasinghe-Dias, K. Feher: "Baseband Pulse Shaping Techniques for Improved Spectral Efficiency $\pi/4$-DQPSK in Non-Linearly Amplified Channels," *Proceedings of the IEEE Vehicular Technology Conference*, VTC-91, St. Louis, MO, May 19-21, 1991.

4.   K. Feher, Ed.: Advanced Digital Communications: Systems and Signal Processing Techniques, Prentice-Hall, Inc., Englewood Cliffs, NJ 07632; 1987.

5.   J.S. Seo, K. Feher: Superposed Quadrature Modulated Baseband Signal Processor, U.S. Patent No. 4,644,565, issued Feb. 17, 1987. Canadian Patent No. 1-265-851, issued Feb. 13, 1990.

6.   K. Feher: "Modulation Techniques for Power and Spectrally Efficient SATCOM Systems," *Proceedings of the Advanced Modulation and Coding Technology Conference*, NASA Lewis Research Center, Cleveland, OH, June 1989.

7.   J.S. Seo, K. Feher: "Bandwidth Compressive 64-State SQAM Modems for Nonlinearly Amplified SATCOM Systems," *International Journal of Satellite Communications*, Vol. 9, July 1991.

FLEXIBLE DIGITAL MODULATION AND CODING SYNTHESIS
FOR SATELLITE COMMUNICATIONS

Mark Vanderaar[*]
Sverdrup Technology, Inc.
2001 Aerospace Parkway
Brookpark, Ohio 44142

James Budinger
NASA Lewis Research Center
Mail Stop 54-8
21000 Brookpark Rd.
Cleveland, Ohio 44135

Craig Hoerig and Dr. John Tague[**]
Dept. of Electrical and Computer
Engineering
Ohio University
Athens, Ohio 45701

ABSTRACT

This paper presents an architecture and a hardware prototype of a Flexible Trellis Modem / Codec (FTMC) transmitter. The theory of operation is built upon a pragmatic approach to trellis-coded modulation that emphasizes power and spectral efficiency [1]. The system incorporates programmable modulation formats, variations of trellis-coding, digital baseband pulse-shaping, and digital channel precompensation. The modulation formats examined include (uncoded and coded) Binary Phase Shift Keying (BPSK), Quartenary Phase Shift Keying (QPSK) , Octal Phase Shift Keying (8PSK), 16-ary Quadrature Amplitude Modulation (16-QAM), and Quadrature Quadrature Phase Shift Keying (Q^{2}PSK) at programmable rates up to 20 Megabits per second (Mbps). The FTMC is part of a developing test bed to quantify modulation and coding concepts.

INTRODUCTION

Regenerative transponders employing onboard-processing and switching techniques are active topics in satellite communications [2, 3, 4]. One common feature among the architectures proposed is the presence of modems and codecs both on the ground and in the spacecraft. The modems and codecs are used with multiple access schemes designed to optimize both uplink and downlink capacities. Generally, they trade some mix of bandwidth and power efficiency.

Frequency Division Multiple Access (FDMA) is an appropriate multiple access technique for low-rate bandwidth efficient uplinks from a large number of ground terminals. In this type of uplink, bandwidth efficiency is directly related to the amount of uplink traffic attainable. Time Division Multiplexing (TDM) is envisioned in high-rate downlinks for its power efficiency. In TDM, nonlinear High Power Amplifiers (HPA's) can be operated at saturation without the adverse effect of intermodulation and distortion as long as the modulated signal maintains a constant envelope. Spread spectrum techniques such as Code Division Multiple Access (CDMA) are used for their power efficiency for the same reasons as TDM. CDMA has also shown increased spectral efficiency in some systems and may be suitable for both uplinks and downlinks.

In the future, as data throughput requirements increase to offer wider band services, TDM or CDM downlinks may not only be limited by available transmit power but also by channel, transponder, and ground terminal bandwidths. For

----

example, Travelling Wave Tube Amplifiers (TWTAs) qualified for space environments today exhibit bandwidths on the order of 1 to 5 percent of the operating frequency [5]. Higher output power TWTAs generally afford lower bandwidth percentages. Technologies such as electronically steerable phased arrays that employ patch antennas can be designed to exhibit bandwidths on the order of 10 percent of their operating frequencies, though they become simpler and smaller at more modest bandwidths of 1 to 3 percent. Also, digital modem and codec hardware can be designed to operate almost exclusively at the symbol rate, as opposed to the bit rate. Thus, higher bit throughput rates can be accomplished at lower processing rate by using modulation schemes with more bits per symbol. This may be useful in a TDM downlink exhibiting high rates to a number of low-cost ground terminals.

Increased flexibility is another dimension in future satellite systems. The on-orbit reconfiguration of the modulation and coding hardware offers the potential to provide more network capacity, increased network availability, multiple service classes, and higher percentage availability at the expense of increased complexity. Efficient use of capacity could be obtained with a network that charges users for the amount of network resources required. Users that require high data and low error rate services would expend more channel bandwidth and/or power per transmitted bit than user with lower quality and rate service requirements. Wider availability may be obtained with flexible transmission formats suited to particular user requirements. Higher percentage availability may be achieved by reconfiguring to more conservative modulation and coding formats overcoming outages such as rain attenuation and multipath fading at the expense of network capacity.

This set of possible operating environments warrant the investigation of modems and codecs that can offer future satellite systems increased operational performance and flexibility.

The first section of this paper presents an overview of the modulation and coding test bed under development. The second section describes the modulation and coding formats supported by the FTMC. The next section develops the digital pulse shaping techniques. Digital baseband precompensation methods for nonlinear channel distortion are then described. The prototype transmitter that includes the baseband modulator and IF upconversion is described next. Finally, conclusions are drawn as to the applicability the techniques presented.

SYSTEM CONFIGURATION

Though this paper focuses on the FTMC transmitter, a brief description is given of the test bed under development. The

proposed FTMC test bed architecture is shown in Figure 1. System control is performed through a personal computer that features a graphical user interface for ease of system configuration and operation. FTMC configuration data is generated via object-oriented simulation software on a workstation. The real time data flow through the FTMC begins with either Pseudo-Random Binary Sequence (PRBS) or patterned test data generated by a digital transmission analyzer (DTA). The DTA is clocked with the FTMC modulator hardware at the programmed data rate. TDM formatting is done by the burst controller. The baseband modulated waveform is then generated by the FTMC modulator. Translation to a 70 MHz IF is performed through quadrature upconversion techniques. The data can then either be translated to RF for amplification via a TWTA or directly to a calibrated IF noise combiner. The signal is then downconverted to baseband where it is filtered, demodulated, and decoded by the FTMC demodulator currently under development.

$$m_I(t) = h(t) * s_I(t)$$
$$m_Q(t) = h(t) * s_Q(t)$$

where $s_I(t)$ and $s_Q(t)$ are the information carrying signals. The information carrying signals can be represented as square pulses over the symbol time T_s of an amplitude corresponding to the constellation definition. The function $h(t)$ is the impulse response of the bandlimiting pulse shape and is described in the next section. Convolution is denoted by the * symbol. In the case of Q^2PSK the information carrying signals correspond to the following expressions

$$s_I(t) = a_1(t)*p_1(t) + a_2(t)*p_2(t)$$
$$s_Q(t) = a_3(t)*p_1(t) + a_4(t)*p_2(t)$$
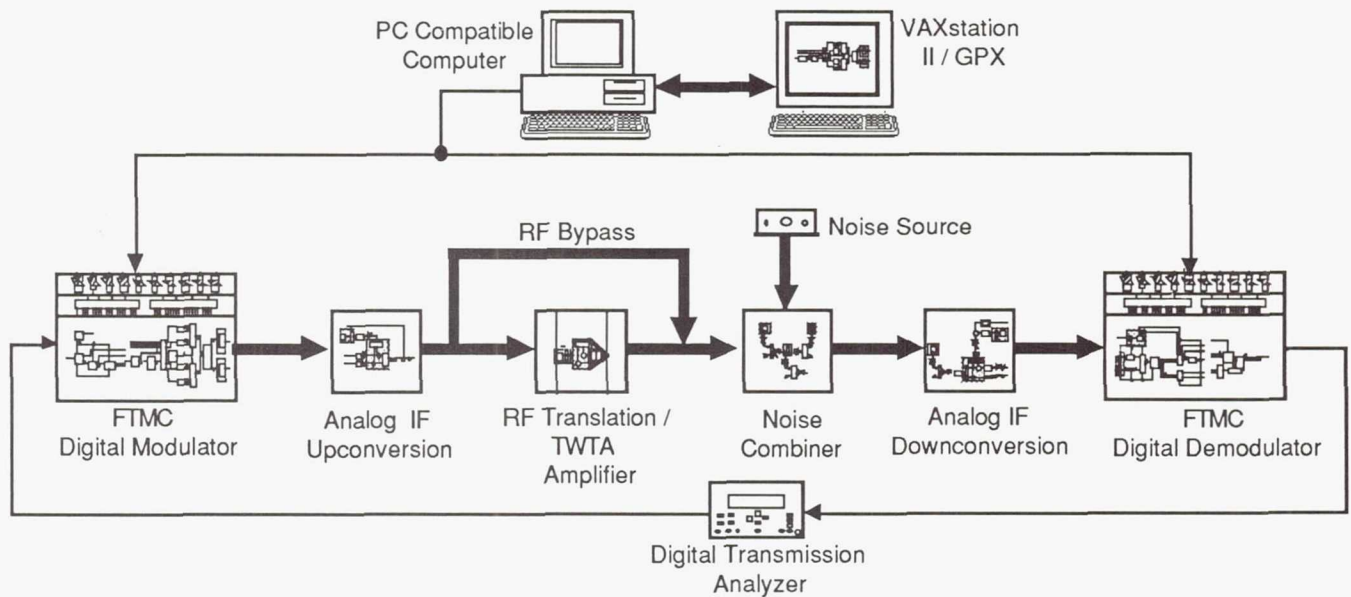
Figure 1. FTMC Test Bed

MODULATION AND CODING

The FTMC supports a number of different modulation formats. Each can be represented in quadrature form by the equation

$$x_{tx}(t) = m_I(t)\cos(2\pi f_c t) + m_Q(t)\sin(2\pi f_c t)$$

where $x_{tx}(t)$ denotes the resulting modulation at a carrier frequency f_c. The signals $m_I(t)$ and $m_Q(t)$ represent the baseband modulating functions. Three Phase Shift Keyed (PSK) techniques are supported that include BPSK, QPSK, and 8PSK. An amplitude and phase modulated technique, 16-QAM, as well as a phase and pulse-shape modulated technique Q^2PSK [6], are also supported . In the PSK and QAM cases the baseband modulating functions represent the in-phase and quadrature bandlimited values associated with the signalling constellation. These can be written as

where each $a_n(t)$ corresponds to a positive or negative amplitude, depending on the information sequence. The $p_1(t)$ and $p_2(t)$ are a set of two orthogonal pulse shapes that also maintain orthogonality with the quadrature mixing operation. These conditions can be defined explicitly by the relations

$$\int_{nT_s}^{(n+1)T_s} p_1(t)p_2(t)dt = 0$$

$$\int_{nT_s}^{(n+1)T_s} p_i(t)e^{j2\pi f_c t}dt = 0, \qquad i=1,2$$

$$f_c = \left(\frac{m}{2T_s}\right), \quad m = \text{integer} \geq 1$$

where n is an integer to define integration over any symbol time. There are an infinite number of $p_i(t)$'s that satisfy the above conditions. A few are developed in [6] that address the issue of phase continuity as well as spectral occupancy. In general, any modulation format whose $S_I(t)$ and $S_Q(t)$ can be represented by two or fewer bits can be created by the FTMC modulator. This is clarified in the waveform synthesis section.

The bit to symbol mapping of all the modulation formats are designed to operate in conjunction with a standard convolutional code of constraint length K=7 and rate 1/2 with generator polynomials $G_{11}=133_8$ and $G_{12}=171_8$ exhibiting the maximal attainable d_{free} of 10 (Figure 2). This approach allows the use of a single encoder and decoder to be used with the programmable modulation formats to achieve a 3 to 7 dB coding gains at real spectral efficiencies (using 40% Nyquist pulse-shaping filters) up to 2.14 bits/second/Hz. Coding gain can be approximated accurately at lower bit-error-rates by the Asymptotic Coding Gain (ACG). The calculation of ACG is based on the most likely decoding error event. The most likely decoding error event for unmerged paths with length greater than one is called the Unmerged Squared Euclidean Distance (USED). As shown in [1] USED for MPSK can be expressed as

$$\text{USED} \geq 2\sin^2\left(\frac{2\pi}{M}\right) + (d_{free} - 4)\sin^2\left(\frac{\pi}{M}\right), \quad M \geq 4$$

where M is the modulation order of the coded scheme. Codes with parallel branches (and thus single branch errors events) have a Parallel Squared Euclidean Distance (PSED), again for MSPK, expressed as

$$\text{PSED} \geq \sin^2\left(\frac{4\pi}{M}\right), \quad M \geq 8$$

ACG is defined as the ratio of the minimum of the USED or PSED to the Normalized Euclidean Distance (NED) for uncoded performance. NED can be written as

$$\text{NED} = \sin^2\left(\frac{2\pi}{M}\right)$$

and thus the ACG (in dB) for MPSK is

$$\text{ACG} \geq 10\log_{10}\left\{4 \text{ Min}\left[\cos^2\left(\frac{2\pi}{M}\right), \frac{1}{2} + \frac{3}{8\cos^2\left(\frac{\pi}{M}\right)}\right]\right\}$$

Evaluations of this expression are given in Table 1. Uncoded performance was calculated using standard bounds [7, 8] for gray coded mappings. Note that for M=4 ACG is determined by the USED and for M=8 ACG is determined by the PSED. A similar argument will be discussed to evaluate QAM performance.

Modulation Format	Code Rate(s)	Asymptotic Coding Gain	~Eb/No for Pb = 10^{-6}	Spectral Efficiency*
BPSK	NA	NA	10.5 dB	0.72
QPSK	NA	NA	10.5 dB	1.43
8PSK	NA	NA	14.0 dB	2.14
16QAM	NA	NA	14.4 dB	2.86
TQPSK	1/2	7.0 dB	4.5 dB	0.72
T8PSK	2/3	3.0 dB	7.5 dB	1.43
T16QAM	2/4	4.3 dB	6.5 dB	1.43
T16QAM	3/4	3.6 dB	10.4 dB	2.14

* In bits/sec/Hz assuming 40% Nyquist Filtering

Table 1 : Modulation and Coding Performance

Figure 2 : Rate 1/2, k=7 Convolutional Encoder

The particulars of each coding scheme can be described by a generator matrix of the form

$$G_{n \to k} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1k} \\ G_{21} & \ddots & \ddots & G_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ G_{n1} & G_{n2} & \cdots & G_{nk} \end{bmatrix}$$

where n is the number of input bits and k is the number of output bits from the encoder. The G_{ij}'s are the activated convolutional encoder shift register taps written in an octal form. The k output bits are mapped into a 2^k- ary modulation scheme in a manner that maximizes the Euclidean distance of the code. For trellis-coded QPSK (TQPSK) the generator matrix is

$$G_{1 \to 2} = \begin{bmatrix} 133_8 & 171_8 \end{bmatrix}$$

Each information bit is coded into one QPSK symbol. A gray mapping of the encoded bits is used. For this code the ACG is determined by the USED of the seven symbol long error event path. The code and mapping structure is illustrated in Figure 3. Note the encoder output bits read from the top to bottom correspond to the bit to symbol assignments read from left to right.
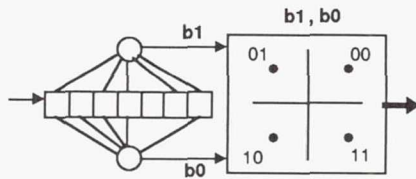
297

Figure 3. Trellis Coded QPSK

Similarly , for T8PSK and T16QAM the generator matrices are

$$G_{2\rightarrow3} = \begin{bmatrix} 133_8 & 171_8 & 0 \\ 0 & 0 & 100_8 \end{bmatrix}$$

$$G_{3\rightarrow4} = \begin{bmatrix} 133_8 & 171_8 & 0 & 0 \\ 0 & 0 & 100_8 & 0 \\ 0 & 0 & 0 & 100_8 \end{bmatrix}$$

The code and mapping structures are shown in Figures 4 and 5.
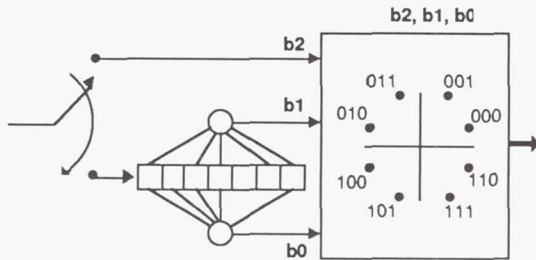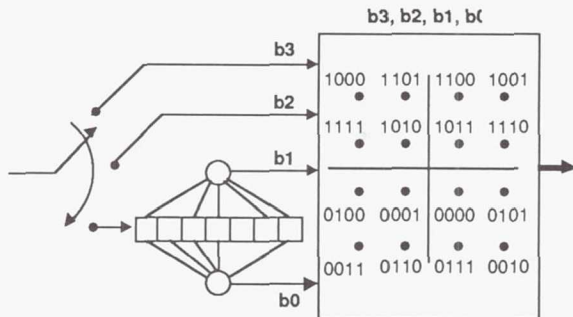

Figure 4: Trellis Coded 8PSK

Figure 5 : Trellis Coded 16-QAM (Mapping #1)

Additionally, a rate 2/4 time-varying code is supported for the 16QAM modulation format. This code maps two information bits into one of 16 modulation points. The first bit is encoded into a two bit symbol that determines the quadrant of the 16-ary symbol. The second information bit is encoded into a two bit symbol that chooses a "sector" within the previously chosen quadrant. This sector and quadrant determine a constellation point for the 16QAM format. The generator matrix is written as

$$G_{2\rightarrow4} = \begin{bmatrix} 133_8 & 171_8 \end{bmatrix} \subset \begin{bmatrix} 133_8 & 171_8 \end{bmatrix}$$

where $\subset$ denotes a demultiplexing operation. The constellation mapping and the code structure is illustrated in Figure 6.

Figure 6: Trellis Coded 16 QAM (Mapping #2)

For Mapping #1 the PSED is the dominated error event. However, it is difficult to calculate due to the nature of the decision regions. However, its performance is given in [1] as approximately 10.4 dB E_b/N_o required to obtain a bit-error-rate of 10^{-6}. This is a coding gain of 3.6 dB as compared to gray coded 8PSK.

For Mapping #2 there are no parallel paths so the PSED is not an issue. The approximate USED is 1.34 as calculated by comparing the first error event path with the all zero path in the trellis diagram. The ACG as compared to the NED of QPSK is

$$ACG \geq 10\log_{10}\left[\frac{USED}{NED}\right] \approx 10\log_{10}\left[\frac{1.34}{\sin^2\left(\frac{\pi}{4}\right)}\right] = 4.3 \text{ dB}$$

WAVEFORM SYNTHESIS

To reduce spectral occupancy, a Static Random Access Memory (SRAM) Distributed Arithmetic (DA) technique is used to synthesize precision baseband pulse-shapes [9, 10, 11]. 40 and 20 percent square-root and full Nyquist raised cosine waveforms have thus far been generated at baseband by sampling and storing the appropriate combinations of transmitted data patterns. The full Nyquist pulse patterns were generated for testing purposes only. The data were sequentially generated for all symbol combinations over specified symbol apertures. An aperture is the length of time for a specified number of symbols to occur.

The frequency response of the square root raised cosine function (SRRCF) is defined by

$$H(f) = \begin{cases} \dfrac{\pi T_s}{\pi(1-\beta)+4\beta}, & 0 \leq \left|\dfrac{f}{f_h}\right| \leq 1-\beta \\[3ex] \dfrac{\pi T_s}{\pi(1-\beta)+4\beta}\cos\left[\dfrac{\pi}{4\beta}\left(\left|\dfrac{f}{f_h}\right|-1+\beta\right)\right], & 1-\beta \leq \left|\dfrac{f}{f_h}\right| \leq 1+\beta \\[3ex] 0, & 1+\beta \leq \left|\dfrac{f}{f_h}\right| \end{cases}$$

where T_s is the symbol period in seconds, ß is the rolloff rate, and f_h is the half-amplitude frequency which equals $1/2T_s$. Taking the inverse Fourier transform of equation H(f), the impulse response is found to be

$$h(t) = \left[\frac{4\beta}{\pi(1-\beta)+4\beta}\right]\left[\frac{T_s \sin\left[\frac{\pi(1-\beta)t}{T_s}\right]}{4\beta t} + \frac{\cos\left[\frac{\pi(1+\beta)t}{T_s}\right]+\frac{4\beta t}{T_s^3}\sin\left[\frac{\pi(1-\beta)t}{T_s}\right]}{1-\left(\frac{4\beta t}{T_s}\right)^2}\right]$$

where h(0) = 1 and

$$h\left(\pm\frac{T_s}{4\beta}\right) = \left[\frac{4\beta}{\pi(1-\beta)+4\beta}\right]\left[\frac{\sin\left[\frac{\pi}{4\beta}(1-\beta)t\right]}{2} + \frac{\pi\left\{\cos\left[\frac{\pi}{4\beta}\right]+\sin\left[\frac{\pi}{4\beta}\right]\right\}}{4\sqrt{2}}\right]$$

The rolloff rate ß is expressed as a percentage and controls the trade-off between the bandwidth and pulse duration. A lower rolloff rate will allow for a narrower passband, but its impulse response will cause more ISI.

Aperture lengths of 6 and 12 symbols were investigated depending on the modulation scheme involved. Data were collected for 8, 16, and 32 samples per symbol. Each symbol was shaped with the SRRCF and the effects of adjacent pulse responses were accounted for over the center symbol period. As the aperture length is an even number, the actual data collected represent the transition from one symbol to the next. The amount of memory required to represent the full set of data on each orthogonal channel for a given modulation format is

memory = sps × levels^{aperture}

where sps is the number of samples per symbol and levels is the number of amplitude levels on one axis of the constellation. The amount of memory can be reduced by exploiting the symmetry in certain modulation formats. The flexible aspect of this pulse shaping procedure is that data can be generated for any rolloff rate, samples per symbol, or symbol aperture length with high precision. A typical software generated eye-diagram is illustrated in Figure 7.

Figure 7: 8PSK Example Eye Diagram.

BASEBAND PRECOMPENSATION

Everything accomplished thus far assumes all subsequent signal transformations are ideal linear operations. However, the nonlinear responses of devices such as TWTA's distort and degrade the signalling scheme. As a second order approximation, a narrow band, frequency independent, memoryless model of a TWTA has Amplitude to Amplitude (AM/AM) and Amplitude to Phase (AM/PM) characteristics that can be modeled as

$$A(r(t)) = \frac{2r(t)}{1+r^2(t)}$$

$$\Phi(r(t)) = \Phi_0\frac{r^2(t)}{1+r^2(t)}$$

where r(t) is the input amplitude level to the TWTA. Many techniques such as precompensation, equalization, and ISI cancelling have been proposed to minimize this type of nonlinearity [12, 13, 14]. In general, transmitter precompensation can occur at baseband, IF or RF. IF and RF precompensation have been applied in many terrestrial microwave radio systems. The limitations observed in these applications are lack of flexibility, long term stability, and difficulty of accurate construction. For these reasons, there has been interest in applying digital signal processing techniques to perform baseband precompensation. Amplitude and phase distortion for an operational TWTA can be determined empirically and thus precisely precompensated. Most implementations thus far have focussed on memoryless systems that warp the signalling constellation with an inverse transform of the amplitude and phase response of the TWTA nonlinearity. However, in systems using pulse shaping, the r becomes r(t) causing the warping to indtroduce additional ISI that results in degraded bit-error-rate performance and spectral regrowth. This problem can be solved by using a memory based architecture [15] similar to the pulse shaping technique. This technique reduces ISI accurately to the symbol aperture. It is interesting to note that this technique is well suited for satellite communications that use lower order modulation schemes. In terrestrial microwave systems, high order modulation schemes are typical. This puts a significant limitation on the aperture and thus accuracy that can be obtained. The memory size as shown in the waveform synthesis section depends exponentially on the number of discrete amplitude levels of the constellation.

The precompensation is implemented in terms of signal space vectors relative to each modulation scheme. The vectors are the complex representation of the in-phase and quadrature sample values. The magnitude and phase of these vectors is precompensated for a set a data that represents the AM/AM and AM/PM characteristics of the nonlinearity. The magnitude precompensation can be accomplished by extending the linear portion of the AM/AM curve. The input signal value enters the precompensator, and the output power is found relative to the linear extension of the data. This value is then found on the actual data curve, and the corrected input backoff corresponding to this new point is output to the rest of the system. Next, the corrected input backoff is further corrected for the AM/PM characteristics of the curve. The phase shift corresponding to this input backoff is found, and this value is subtracted from the phase of the signal entering the TWTA. An example eye diagram of precompensated data for BPSK is shown in Figure 8. A square root 40 percent raised cosine pulse was used to shape the data at 16 samples per symbol with a twleve symbol aperture. Measured AM/AM and AM/PM curves from a TWTA were used to model the nonlinearity. Figure 10 shows the precompensated scatter plot. As shown in Figure 9 the precompensated data exhibits increased spectral occupancy that must be included in the bandwidth calculations for the modulation upconversion and amplification functions. The clustering and ISI effects seen in the eye diagram are products of the square root filter. A full raised cosine or matched square root raised cosines do not exhibit these effects. Figure 11 shows the eye diagram after the TWTA distortion and a matched receive finite impulse response filter with 96 taps.

This type of precompensation would be particularly useful in an operational system. A network using precompensation could adapt and load new precompensation data as nonlinear channel characteristics changed. Comparisons are being generated that compare the total degradation for a set of modulation formats as a function of input backoff, filtering, and symbol aperture for baseband precompensated signals.

Figure 9: Precompensated spectrum.

Figure 10: Precompensated scatter plot.

Figure 11: Receive filter output eye diagram.

FTMC TRANSMITTER HARDWARE

Prototype hardware has been constructed to verify FTMC concepts with real system nonidealities. The transmitter block diagram is shown in Figure 12 and the FTMC modem chassis layout in Figure 13. Serial data can be clocked in at TTL levels continuously or in a bursted format. This data is then translated to a parallel format of one to four bits wide. One bit is supplied to the rate 1/2 k=7 convolutional encoder, and the rest bypass it. Uncoded, coded, and burst control bits (total of nine address bits) are supplied to a modulation

Figure 8: Precompensated eye diagram.

mapping function. This modulation mapper is a 512 X 4 bit SRAM loaded with four bits of information that represent four in-phase and four quadrature levels. This information (either data or preamble) is turned broadside to generate symbol apertures via shift registers. Symbol apertures are transformed into pulse-shaped symbols with the Nyquist pattern generators. The pattern generators are each 128K X 8 SRAM's. The SRAM's are loaded with the patterns generated via the techniques discussed in the waveform synthesis section. The eight bit samples output from the filter SRAMs are either directly output to the digital to analog converters or used as address bit to the precompensator. The precompensator performs the nonlinear transformation as described in the precompensation section. The digital to analog converters use eight bit quantization. Figures 14 through 19 show eye diagrams and spectrums of four of the modulation formats, using full 40 percent raised cosine pulse shapes. In each figure, the sample rate is 32.768 MHz and there are 32 samples per symbol. This gives a bit rate of

$$\text{Bit rate} = \frac{\text{sample rate}}{\text{samples per symbol}} \log_2(M))$$

Each spectrum is plotted in a frequency window corresponding to the width of the main lobe, $2/T_b$, of a BPSK signal with the same bit rate, where T_b is the bit time.

Figure 13. FMC Chassis Mechanical Layout

Figure 12 : Digital Baseband Modulator / Encoder

The IF upconversion module translates the baseband modulated signal to 70 MHz using standard quadrature upconversion techniques (Figure 20). The 70 MHz carrier is obtained from an analog signal generator that can be phase locked to the symbol time to satisfy Q^2PSK orthogonality conditions. Direct Digital Synthesizers (DDS) could also be used to generate the IF carrier.

The quadrature spectral reconstruction filters are designed to accommodate the full range of FTMC operation in terms of data and sampling rate. The signal bandwidth (assuming no spectral shaping) is obtained as

$$BW = \frac{f_s}{sps}$$

where f_s is the sampling frequency and sps is the number of samples per symbol. The maximum bandwidth, BW_{max} is 5

MHz. This occurs when f_s is 40 MHz and sps is 8 samples per symbol. The spectral reconstruction filters must contain this frequency in their passband. The other constraint defining the frequency response of the reconstruction filters is the lowest alias frequency. This occurs at $f_s = 10$ MHz and sps = 8, giving

$$f_{alias} = f_s - \frac{f_s}{sps} = 10 \text{ MHz} - \frac{10 \text{ MHz}}{8 \text{ sam/sym}} = 8.75 \text{ MHz}$$

This f_{alias} of 8.75 MHz must be in the stopband of the filter response. The resulting filter mask is shown in Figure 21, with the 48 dB of attenuation specification obtained from the quantization error of the 8 bit digital to analog converters. Note that the eventual limit of available sample rates and data rates is limited by the lack of flexibility in the reconstruction filter. To maintain the low level of ISI obtained by the Nyquist filters the group delay variation is limited to
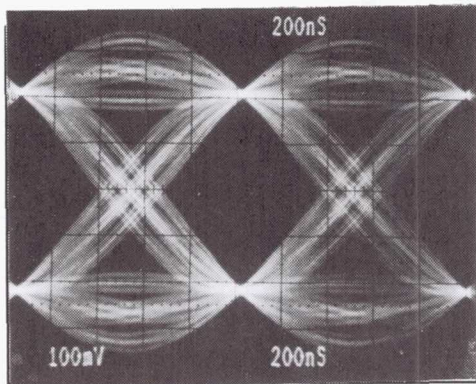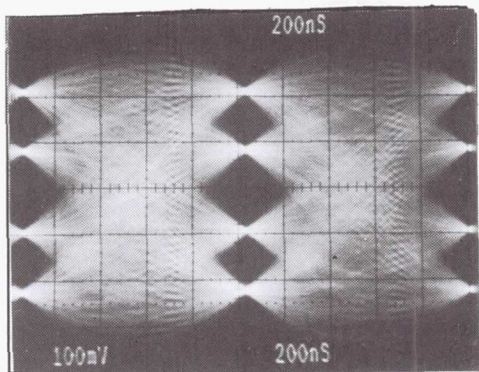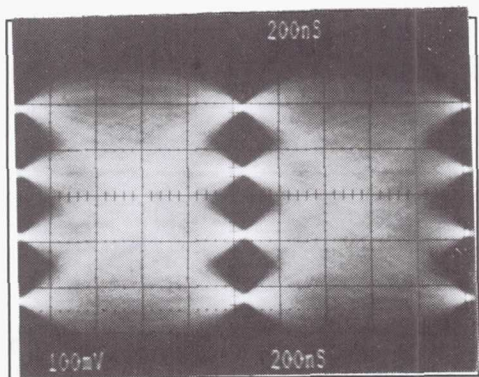
301

Figure14: BPSK Eye Diagram

Figure 15: BPSK Spectrum

Figure16: 8PSK Eye Diagram

Figure 17: 8PSK Spectrum

Figure18: 16 QAM Eye Diagram
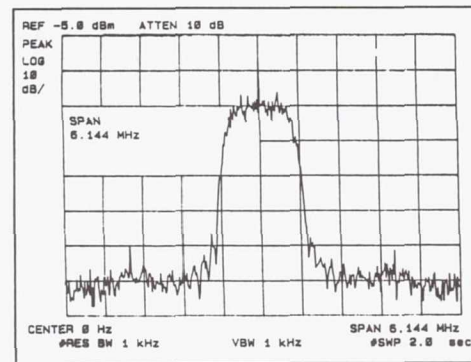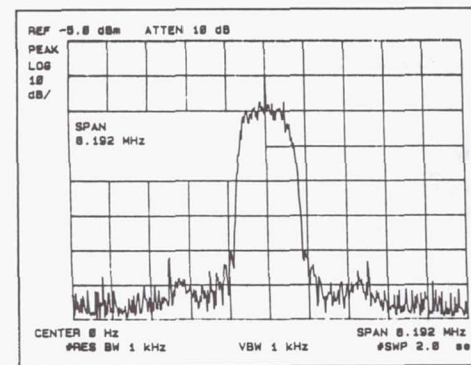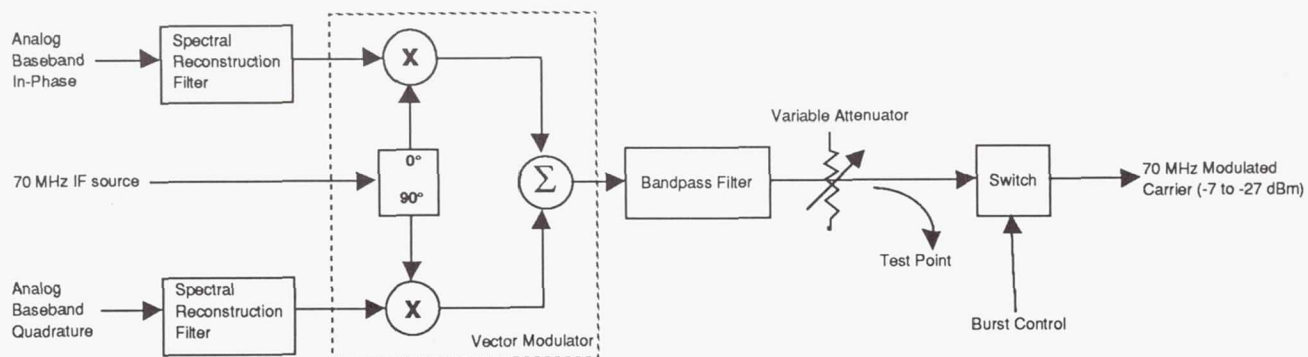

Figure 19: 16 QAM Spectrum

Figure 20 : Analog IF Upconversion

15 nanoseconds up to 3.5 MHz. This means that 40 percent Nyquist filters (cutoff approximately 3.5 MHz) will have a group delay over their entire bandwidth of no more than

$$\tau_{max} = \frac{\tau_{max\ symbol\ BW}}{T_{symbol\ max}} = \frac{15\ nS}{200\ nS} = 7.5\ \%$$

for the worst case symbol rate of 5 Msps. This effect causes an expected E_b/N_0 degradation of no more than a few tenths of a dB.

Figure 21 : Spectral Reconstruction Filter Mask

CONCLUSION

The architecture and hardware prototype of a flexible trellis-coded modem and codec transmitter has been presented. The FTMC transmitter uses digital modulation and coding, multi-symbol digital pulse shaping, and digital precompensation to tradeoff complexity for improved signal quality, bandwidth, and power efficiency. This type of architecture is envisioned to be of use in FDMA satellite uplinks and TDM or CDM satellite downlinks.

In any planned operational satellite network, the choice of modulation and coding techniques is dictated heavily by numerous other system concerns. However, offering future system designers workable flexible modems and codecs may relieve specifications in other subsystems and offer an enhanced satellite system that can offer more varied and higher quality services.

ACKNOWLEDGEMENT

REFERENCES

[1] Viterbi, A. J., et al: A Pragmatic Approach to Trellis-Coded Modulation. IEEE Communications Magazine, July 1989, pp. 11-19.

[2] Harrold, J. L.; Budinger, J. M.; and Stevens,G. H.: On-Board Switching and Processing, Proc. IEEE, vol. 78, No. 7, July 1990, pp.1206-1212.

[3] Nuspl, P. P., et al: On-board Processing for Communications Satellites : Systems and Benefits, International Journal of Satellite Communications, Vol. 5, 1987, pp. 65-76.

[4] Campanella, S. J.: Communications Satellites: Orbiting into the 90's, IEEE Spectrum Magazine, August 1990, pp. 49-52.

[5] Hansen, J. W., et el: US TWTs from 1 to 100 GHz, Microwave Journal • 1989 State of the Art Reference, pp. 179- 193.

[6] Saha, D.; and Birdsall, T. G.: Quadrature-Quadrature Phase-Shift Keying, IEEE Transactions on Communications, Vol. 37, No. 5, May 1990, pp. 437- 448.

[7] Sklar, B.: Digital Communications Fundamentals and Applications, Prentice Hall, 1988.

[8] Lee, P. J.: Computation of the Bit Error Rate of Coherent M-ary PSK with Gray Code Bit Mapping, IEEE Transactions on Communication, Vol. 34, No. 5, May 1986, pp 488- 491.

[9] Poklemba, J. J: Programmable Digital Modem, NASA Space Communications Technology Conference, November 1991.

[10] Aghvami, A. H.; and Gemikonakli, O.: 16-ary QAM System for Low Data-Rate Satellite Services, Electron. Lett., Vol. 25, No. 16, August 3, 1989, pp. 1055- 1056.

[11] Allan, R. D.; Bramwell, J. R.; and Tomlinson, M.: A High Performance Satellite Data Modem Using Real-Time Digital Signal Processing. Olympus Utilisation Conference, Vienna, Italy, May 1989, pp. 51-55.

[12] Karam, G.; and Sari, H.: Analysis of Predistortion, Equalization, and ISI Cancellation Techniques in Digital Radio Systems with Nonlinear Transmit Amplifiers.

[13] Biglieri, E.; Barberis, S.; and Catena, M.: Analysis and Compensation of Nonlinearities in Digital Transmission Systems, IEEE Journal on Selected Areas in Communications, Vol. 6, No. 1, January 1988, pp. 42- 51.

[14] Cheung, S. W.; Aghvami, A.H.: Performance of a 16-ary DEQAM Modem Employing a baseband or RF Predistorter Over a Regenerative Satellite Link, IEE Proceedings, Vol. 135, No. 6, December 1988, pp. 547- 557.

[15] Karam, G.; and Sari, H.: A Data Predistortion Technique with Memory for QAM Radio Systems, IEEE Transactions on Communications, Vol. 39 , No. 2, February 1991, pp. 336-344.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE November 1991	3. REPORT TYPE AND DATES COVERED Conference Publication

4. TITLE AND SUBTITLE

Space Communications Technology Conference
Onboard Processing and Switching

5. FUNDING NUMBERS

WU-650-60-21

6. AUTHOR(S)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135-3191

8. PERFORMING ORGANIZATION REPORT NUMBER

E-6548

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D.C. 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA CP-3132

11. SUPPLEMENTARY NOTES

Responsible person, James M. Budinger, (216) 433-3496.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Categories 17 and 32

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This proceedings of the NASA Space Communications Technology Conference held on November 12-14, 1991, in Cleveland, Ohio, presents onboard processing and switching technologies for application in future satellite communications systems. Onboard processing for B-ISDN (broadband integrated services digital network) services, packet-switched FDMA/TDM (frequency division multiple access/time division), a novel variation of CDMA (code division multiple access), and the merits of OBP (onboard processing) for low Earth orbiting mobile systems are addressed in the section, Satellite Network Architectures. Telecommunications services for the next generation of networks, a traffic analysis of B-ISDN, advanced lightwave networks, and survivable system processing are covered in Network Control and Protocols. Papers on bandwidth efficient coding, artificial intelligence for Earth terminals, multicarrier demodulation, neural-network-based video compression, an expert system for fault diagnosis, and laboratory measurements of INTELSAT OBP subsystems are presented in Concurrent Poster Presentations and Demonstrations. Two papers on expert systems for space hardware diagnostics and fault tolerance applied to fiber-optic networks and multichannel demultiplexers are covered in Fault Tolerance and Autonomy. Digital, optical, and acousto-optical approaches to multichannel demultiplexing and demodulation for onboard processed FDMA are presented in Multichannel Demultiplexing and Demodulation. In Information Switching and Routing, an onboard baseband switch, an OBP payload for asynchronous relay satellite networks, OBP for future telecommunications services, and congestion control for satellite packet switching are presented. In Modulation and Coding, the final section of this document, advanced modulation and coding technologies are applied to a B-ISDN modem-codec, a flexible, high-speed codec, and two variable-rate digital modems. Decoding multilevel demodulation and two approaches to digital pulse-shaped modulation for nonlinear satellite communications are also presented. Papers for the final two sessions of the conference were not available at the time of publication. These sessions, which addressed planned communications satellite systems under development at INTELSAT, NASA, Motorola, and Space Systems/Loral, and a panel session on the issues and challenges of onboard processing and switching technology insertion, are covered in a companion publication.

14. SUBJECT TERMS

Onboard processing; Modulation; Coding; Autonomy; Switching;
Routing; Network control

15. NUMBER OF PAGES

312

16. PRICE CODE

A14

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT

$$f_c = \left(\frac{m}{2T_s}\right), \quad m = \text{integer} \geq 1$$

where n is an integer to define integration over any symbol time. There are an infinite number of $p_i(t)$'s that satisfy the above conditions. A few are developed in [6] that address the issue of phase continuity as well as spectral occupancy. In general, any modulation format whose $S_I(t)$ and $S_Q(t)$ can be represented by two or fewer bits can be created by the FTMC modulator. This is clarified in the waveform synthesis section.

The bit to symbol mapping of all the modulation formats are designed to operate in conjunction with a standard convolutional code of constraint length K=7 and rate 1/2 with generator polynomials $G_{11}=133_8$ and $G_{12}=171_8$ exhibiting the maximal attainable d_{free} of 10 (Figure 2). This approach allows the use of a single encoder and decoder to be used with the programmable modulation formats to achieve a 3 to 7 dB coding gains at real spectral efficiencies (using 40% Nyquist pulse-shaping filters) up to 2.14 bits/second/Hz. Coding gain can be approximated accurately at lower bit-error-rates by the Asymptotic Coding Gain (ACG). The calculation of ACG is based on the most likely decoding error event. The most likely decoding error event for unmerged paths with length greater than one is called the Unmerged Squared Euclidean Distance (USED). As shown in [1] USED for MPSK can be expressed as

$$USED \geq 2\sin^2\left(\frac{2\pi}{M}\right) + (d_{free}- 4)\sin^2\left(\frac{\pi}{M}\right), \quad M \geq 4$$

where M is the modulation order of the coded scheme. Codes with parallel branches (and thus single branch errors events) have a Parallel Squared Euclidean Distance (PSED), again for MSPK, expressed as

$$PSED \geq \sin^2\left(\frac{4\pi}{M}\right), \quad M \geq 8$$

ACG is defined as the ratio of the minimum of the USED or PSED to the Normalized Euclidean Distance (NED) for uncoded performance. NED can be written as

$$NED = \sin^2\left(\frac{2\pi}{M}\right)$$

and thus the ACG (in dB) for MPSK is

$$ACG \geq 10\log_{10}\left\{4 \ Min\left[\cos^2\left(\frac{2\pi}{M}\right), \ \frac{1}{2} + \frac{3}{8\cos^2\left(\frac{\pi}{M}\right)}\right]\right\}$$

Evaluations of this expression are given in Table 1. Uncoded performance was calculated using standard bounds [7, 8] for gray coded mappings. Note that for M=4 ACG is determined by the USED and for M=8 ACG is determined by the PSED. A similar argument will be discussed to evaluate QAM performance.

Modulation Format	Code Rate(s)	Asymptotic Coding Gain	~Eb/No for Pb = 10^-6	Spectral Efficiency*
BPSK	NA	NA	10.5 dB	0.72
QPSK	NA	NA	10.5 dB	1.43
8PSK	NA	NA	14.0 dB	2.14
16QAM	NA	NA	14.4 dB	2.86
TQPSK	1/2	7.0 dB	4.5 dB	0.72
T8PSK	2/3	3.0 dB	7.5 dB	1.43
T16QAM	2/4	4.3 dB	6.5 dB	1.43
T16QAM	3/4	3.6 dB	10.4 dB	2.14

* In bits/sec/Hz assuming 40% Nyquist Filtering
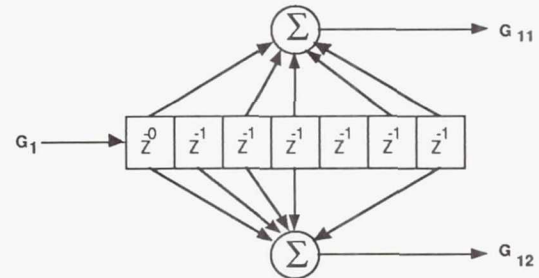Table 1 : Modulation and Coding Performance

Figure 2 : Rate 1/2, k=7 Convolutional Encoder

The particulars of each coding scheme can be described by a generator matrix of the form

$$G_{n \to k} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1k} \\ G_{21} & \ddots & \ddots & G_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ G_{n1} & G_{n2} & \cdots & G_{nk} \end{bmatrix}$$

where n is the number of input bits and k is the number of output bits from the encoder. The G_{ij}'s are the activated convolutional encoder shift register taps written in an octal form. The k output bits are mapped into a 2^k- ary modulation scheme in a manner that maximizes the Euclidean distance of the code. For trellis-coded QPSK (TQPSK) the generator matrix is

$$G_{1 \to 2} = \begin{bmatrix} 133_8 & 171_8 \end{bmatrix}$$

Each information bit is coded into one QPSK symbol. A gray mapping of the encoded bits is used. For this code the ACG is determined by the USED of the seven symbol long error event path. The code and mapping structure is illustrated in Figure 3. Note the encoder output bits read from the top to bottom correspond to the bit to symbol assignments read from left to right.
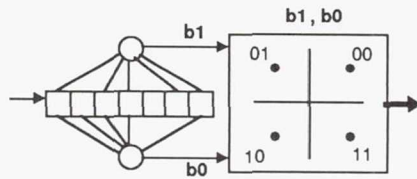
Figure 3. Trellis Coded QPSK

Similarly , for T8PSK and T16QAM the generator matrices are

$$G_{2\to3} = \begin{bmatrix} 133_8 & 171_8 & 0 \\ 0 & 0 & 100_8 \end{bmatrix}$$

$$G_{3\to4} = \begin{bmatrix} 133_8 & 171_8 & 0 & 0 \\ 0 & 0 & 100_8 & 0 \\ 0 & 0 & 0 & 100_8 \end{bmatrix}$$

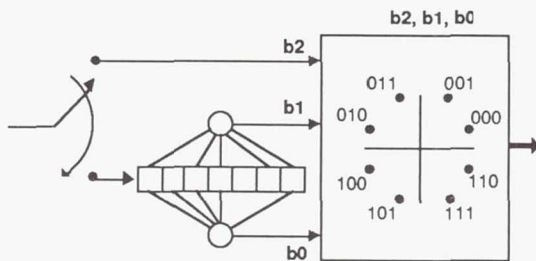The code and mapping structures are shown in Figures 4 and 5.
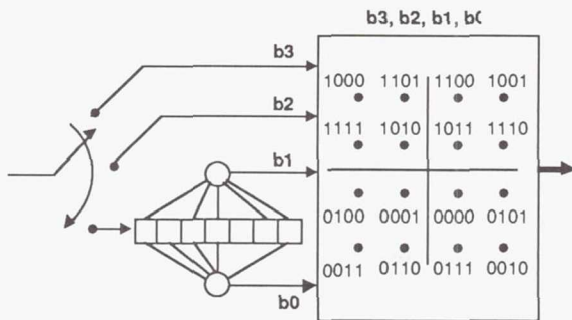

Figure 4: Trellis Coded 8PSK

Figure 5 : Trellis Coded 16-QAM (Mapping #1)

Additionally, a rate 2/4 time-varying code is supported for the 16QAM modulation format. This code maps two information bits into one of 16 modulation points. The first bit is encoded into a two bit symbol that determines the quadrant of the 16-ary symbol. The second information bit is encoded into a two bit symbol that chooses a "sector" within the previously chosen quadrant. This sector and quadrant determine a constellation point for the 16QAM format. The generator matrix is written as

$$G_{2\to4} = \begin{bmatrix} 133_8 & 171_8 \end{bmatrix} \subset \begin{bmatrix} 133_8 & 171_8 \end{bmatrix}$$

where $\subset$ denotes a demultiplexing operation. The constellation mapping and the code structure is illustrated in Figure 6.

Figure 6: Trellis Coded 16 QAM (Mapping #2)

For Mapping #1 the PSED is the dominated error event. However, it is difficult to calculate due to the nature of the decision regions. However, its performance is given in [1] as approximately 10.4 dB E_b/N_o required to obtain a bit-error-rate of 10^{-6}. This is a coding gain of 3.6 dB as compared to gray coded 8PSK.

For Mapping #2 there are no parallel paths so the PSED is not an issue. The approximate USED is 1.34 as calculated by comparing the first error event path with the all zero path in the trellis diagram. The ACG as compared to the NED of QPSK is

$$ACG \geq 10\log_{10}\left[\frac{USED}{NED}\right] \approx 10\log_{10}\left[\frac{1.34}{\sin^2\left(\frac{\pi}{4}\right)}\right] = 4.3 \text{ dB}$$

WAVEFORM SYNTHESIS

To reduce spectral occupancy, a Static Random Access Memory (SRAM) Distributed Arithmetic (DA) technique is used to synthesize precision baseband pulse-shapes [9, 10, 11]. 40 and 20 percent square-root and full Nyquist raised cosine waveforms have thus far been generated at baseband by sampling and storing the appropriate combinations of transmitted data patterns. The full Nyquist pulse patterns were generated for testing purposes only. The data were sequentially generated for all symbol combinations over specified symbol apertures. An aperture is the length of time for a specified number of symbols to occur.

The frequency response of the square root raised cosine function (SRRCF) is defined by

$$H(f) = \begin{cases} \dfrac{\pi T_s}{\pi(1-\beta)+4\beta}, & 0 \leq \left|\dfrac{f}{f_h}\right| \leq 1-\beta \\[2ex] \dfrac{\pi T_s}{\pi(1-\beta)+4\beta}\cos\left[\dfrac{\pi}{4\beta}\left(\left|\dfrac{f}{f_h}\right|-1+\beta\right)\right], & 1-\beta \leq \left|\dfrac{f}{f_h}\right| \leq 1+\beta \\[2ex] 0, & 1+\beta \leq \left|\dfrac{f}{f_h}\right| \end{cases}$$

where T_s is the symbol period in seconds, β is the rolloff rate, and f_h is the half-amplitude frequency which equals $1/2T_s$. Taking the inverse Fourier transform of equation H(f), the impulse response is found to be

298

$$h(t) = \left[\frac{4\beta}{\pi(1-\beta)+4\beta}\right]\left[\frac{T_s \sin\left[\frac{\pi(1-\beta)t}{T_s}\right]}{4\beta t} + \frac{\cos\left[\frac{\pi(1+\beta)t}{T_s}\right]+\frac{4\beta t}{T_s^3}\sin\left[\frac{\pi(1-\beta)t}{T_s}\right]}{1-\left(\frac{4\beta t}{T_s}\right)^2}\right]$$

where $h(0) = 1$ and

$$h\left(\pm\frac{T_s}{4\beta}\right) = \left[\frac{4\beta}{\pi(1-\beta)+4\beta}\right]\left[\frac{\sin\left[\frac{\pi}{4\beta}(1-\beta)t\right]}{2} + \frac{\pi\left\{\cos\left[\frac{\pi}{4\beta}\right]+\sin\left[\frac{\pi}{4\beta}\right]\right\}}{4\sqrt{2}}\right]$$

The rolloff rate ß is expressed as a percentage and controls the trade-off between the bandwidth and pulse duration. A lower rolloff rate will allow for a narrower passband, but its impulse response will cause more ISI.

Aperture lengths of 6 and 12 symbols were investigated depending on the modulation scheme involved. Data were collected for 8, 16, and 32 samples per symbol. Each symbol was shaped with the SRRCF and the effects of adjacent pulse responses were accounted for over the center symbol period. As the aperture length is an even number, the actual data collected represent the transition from one symbol to the next. The amount of memory required to represent the full set of data on each orthogonal channel for a given modulation format is

$$\text{memory} = \text{sps} \times \text{levels}^{\text{aperture}}$$

where sps is the number of samples per symbol and levels is the number of amplitude levels on one axis of the constellation. The amount of memory can be reduced by exploiting the symmetry in certain modulation formats. The flexible aspect of this pulse shaping procedure is that data can be generated for any rolloff rate, samples per symbol, or symbol aperture length with high precision. A typical software generated eye-diagram is illustrated in Figure 7.
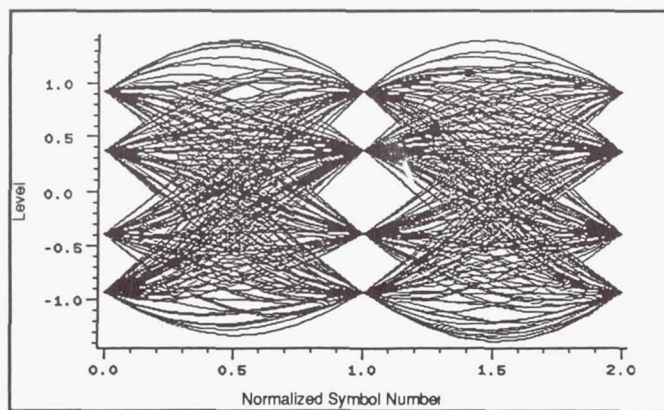

Figure 7: 8PSK Example Eye Diagram.

BASEBAND PRECOMPENSATION

Everything accomplished thus far assumes all subsequent signal transformations are ideal linear operations. However, the nonlinear responses of devices such as TWTA's distort and degrade the signalling scheme. As a second order approximation, a narrow band, frequency independent, memoryless model of a TWTA has Amplitude to Amplitude (AM/AM) and Amplitude to Phase (AM/PM) characteristics that can be modeled as

$$A(r(t)) = \frac{2r(t)}{1+r^2(t)}$$

$$\Phi(r(t)) = \Phi_0\frac{r^2(t)}{1+r^2(t)}$$

where r(t) is the input amplitude level to the TWTA. Many techniques such as precompensation, equalization, and ISI cancelling have been proposed to minimize this type of nonlinearity [12, 13, 14]. In general, transmitter precompensation can occur at baseband , IF or RF. IF and RF precompensation have been applied in many terrestrial microwave radio systems. The limitations observed in these applications are lack of flexibility, long term stability, and difficulty of accurate construction. For these reasons, there has been interest in applying digital signal processing techniques to perform baseband precompensation. Amplitude and phase distortion for an operational TWTA can be determined empirically and thus precisely precompensated. Most implementations thus far have focussed on memoryless systems that warp the signalling constellation with an inverse transform of the amplitude and phase response of the TWTA nonlinearity. However, in systems using pulse shaping, the r becomes r(t) causing the warping to indtroduce additional ISI that results in degraded bit-error-rate performance and spectral regrowth. This problem can be solved by using a memory based architecture [15] similar to the pulse shaping technique. This technique reduces ISI accurately to the symbol aperture. It is interesting to note that this technique is well suited for satellite communications that use lower order modulation schemes. In terrestrial microwave systems, high order modulation schemes are typical. This puts a significant limitation on the aperture and thus accuracy that can be obtained. The memory size as shown in the waveform synthesis section depends exponentially on the number of discrete amplitude levels of the constellation.

299

The precompensation is implemented in terms of signal space vectors relative to each modulation scheme. The vectors are the complex representation of the in-phase and quadrature sample values. The magnitude and phase of these vectors is precompensated for a set a data that represents the AM/AM and AM/PM characteristics of the nonlinearity. The magnitude precompensation can be accomplished by extending the linear portion of the AM/AM curve. The input signal value enters the precompensator, and the output power is found relative to the linear extension of the data. This value is then found on the actual data curve, and the corrected input backoff corresponding to this new point is output to the rest of the system. Next, the corrected input backoff is further corrected for the AM/PM characteristics of the curve. The phase shift corresponding to this input backoff is found, and this value is subtracted from the phase of the signal entering the TWTA. An example eye diagram of precompensated data for BPSK is shown in Figure 8. A square root 40 percent raised cosine pulse was used to shape the data at 16 samples per symbol with a twleve symbol aperture. Measured AM/AM and AM/PM curves from a TWTA were used to model the nonlinearity. Figure 10 shows the precompensated scatter plot. As shown in Figure 9 the precompensated data exhibits increased spectral occupancy that must be included in the bandwidth calculations for the modulation upconversion and amplification functions. The clustering and ISI effects seen in the eye diagram are products of the square root filter. A full raised cosine or matched square root raised cosines do not exhibit these effects. Figure 11 shows the eye diagram after the TWTA distortion and a matched receive finite impulse response filter with 96 taps.
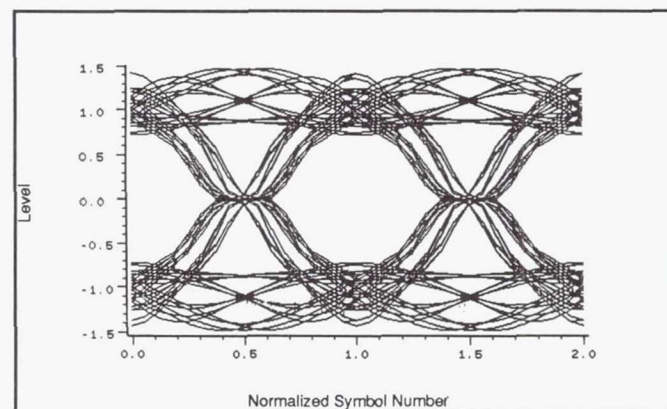
This type of precompensation would be particularly useful in an operational system. A network using precompensation could adapt and load new precompensation data as nonlinear channel characteristics changed. Comparisons are being generated that compare the total degradation for a set of modulation formats as a function of input backoff, filtering, and symbol aperture for baseband precompensated signals.

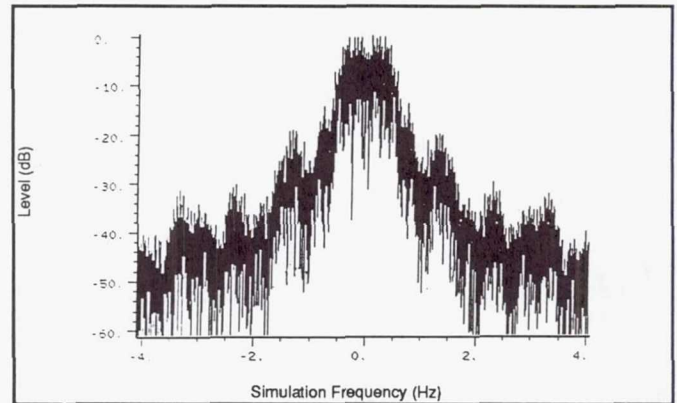Figure 9: Precompensated spectrum.

Figure 10: Precompensated scatter plot.
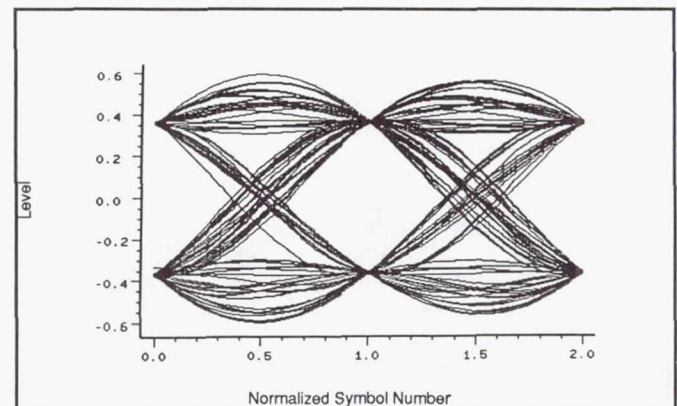

Figure 11: Receive filter output eye diagram.

Figure 8: Precompensated eye diagram.

FTMC TRANSMITTER HARDWARE

Prototype hardware has been constructed to verify FTMC concepts with real system nonidealities. The transmitter block diagram is shown in Figure 12 and the FTMC modem chassis layout in Figure 13. Serial data can be clocked in at TTL levels continuously or in a bursted format. This data is then translated to a parallel format of one to four bits wide. One bit is supplied to the rate 1/2 k=7 convolutional encoder, and the rest bypass it. Uncoded, coded, and burst control bits (total of nine address bits) are supplied to a modulation

mapping function. This modulation mapper is a 512 X 4 bit SRAM loaded with four bits of information that represent four in-phase and four quadrature levels. This information (either data or preamble) is turned broadside to generate symbol apertures via shift registers. Symbol apertures are transformed into pulse-shaped symbols with the Nyquist pattern generators. The pattern generators are each 128K X 8 SRAM's. The SRAM's are loaded with the patterns generated via the techniques discussed in the waveform synthesis section. The eight bit samples output from the filter SRAMs are either directly output to the digital to analog converters or used as address bit to the precompensator. The precompensator performs the nonlinear transformation as described in the precompensation section. The digital to analog converters use eight bit quantization. Figures 14 through 19 show eye diagrams and spectrums of four of the modulation formats, using full 40 percent raised cosine pulse shapes. In each figure, the sample rate is 32.768 MHz and there are 32 samples per symbol. This gives a bit rate of

$$\text{Bit rate} = \frac{\text{sample rate}}{\text{samples per symbol}} \log_2(M))$$

Each spectrum is plotted in a frequency window corresponding to the width of the main lobe, $2/T_b$, of a BPSK signal with the same bit rate, where T_b is the bit time.

Figure 13. FMC Chassis Mechanical Layout

Figure 12 : Digital Baseband Modulator / Encoder

The IF upconversion module translates the baseband modulated signal to 70 MHz using standard quadrature upconversion techniques (Figure 20). The 70 MHz carrier is obtained from an analog signal generator that can be phase locked to the symbol time to satisfy Q^2PSK orthogonality conditions. Direct Digital Synthesizers (DDS) could also be used to generate the IF carrier.

The quadrature spectral reconstruction filters are designed to accommodate the full range of FTMC operation in terms of data and sampling rate. The signal bandwidth (assuming no spectral shaping) is obtained as

$$BW = \frac{f_s}{sps}$$

where f_s is the sampling frequency and sps is the number of samples per symbol. The maximum bandwidth, BW_{max} is 5
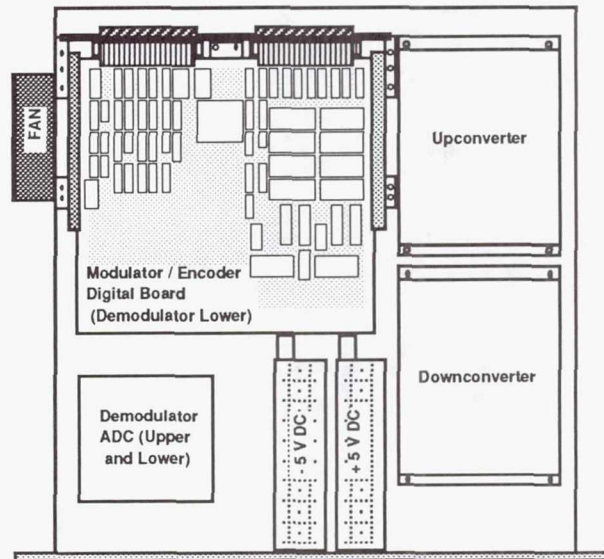
MHz. This occurs when f_s is 40 MHz and sps is 8 samples per symbol. The spectral reconstruction filters must contain this frequency in their passband. The other constraint defining the frequency response of the reconstruction filters is the lowest alias frequency. This occurs at f_s = 10 MHz and sps = 8, giving

$$f_{alias} = f_s - \frac{f_s}{sps} = 10 \text{ MHz} - \frac{10 \text{ MHz}}{8 \text{ sam/sym}} = 8.75 \text{ MHz}$$

This f_{alias} of 8.75 MHz must be in the stopband of the filter response. The resulting filter mask is shown in Figure 21, with the 48 dB of attenuation specification obtained from the quantization error of the 8 bit digital to analog converters. Note that the eventual limit of available sample rates and data rates is limited by the lack of flexibility in the reconstruction filter. To maintain the low level of ISI obtained by the Nyquist filters the group delay variation is limited to
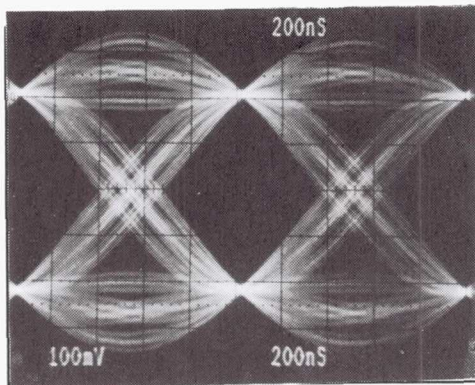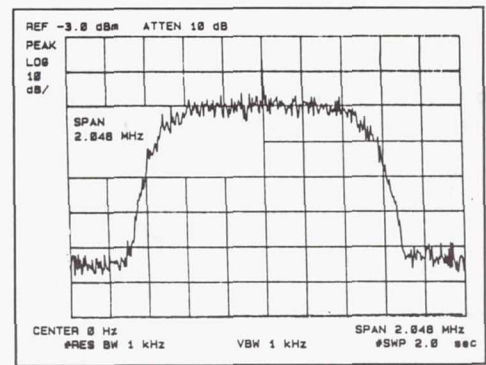
301

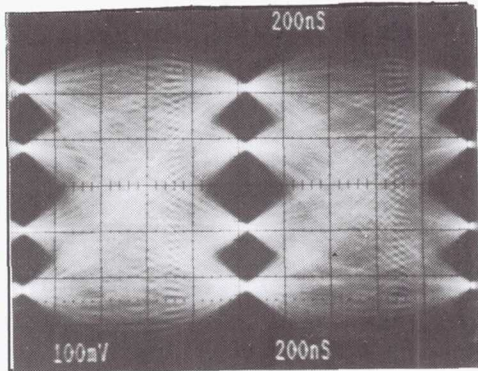Figure14: BPSK Eye Diagram

Figure 15: BPSK Spectrum
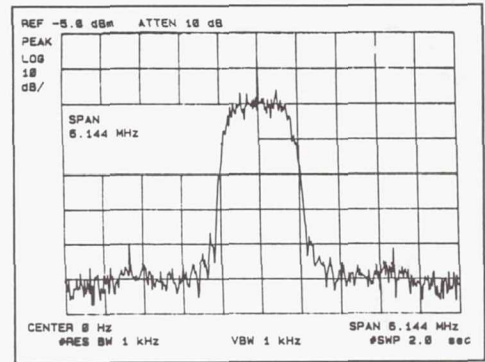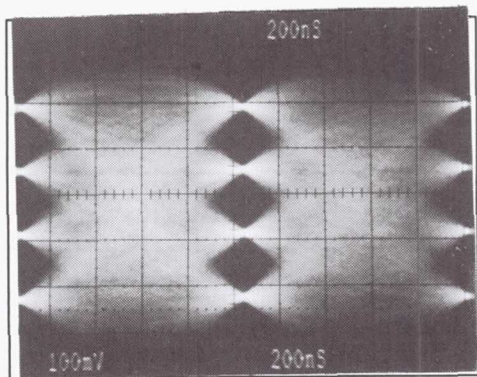

Figure16: 8PSK Eye Diagram

Figure 17: 8PSK Spectrum
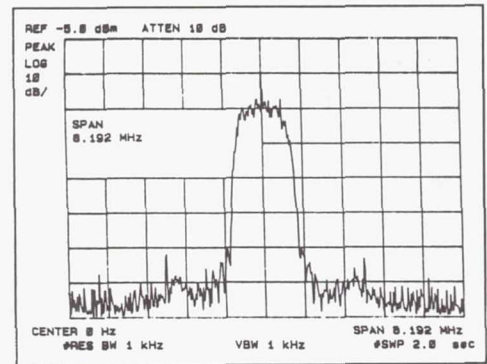

Figure18: 16 QAM Eye Diagram
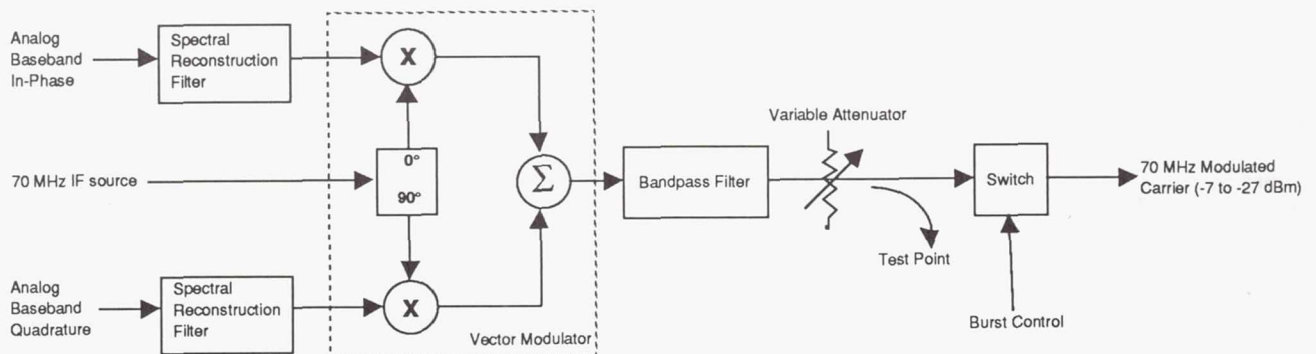

Figure 19: 16 QAM Spectrum

Figure 20 : Analog IF Upconversion

15 nanoseconds up to 3.5 MHz. This means that 40 percent Nyquist filters (cutoff approximately 3.5 MHz) will have a group delay over their entire bandwidth of no more than

$$\tau_{max} = \frac{\tau_{max\ symbol\ BW}}{T_{symbol\ max}} = \frac{15\ nS}{200\ nS} = 7.5\ \%$$

for the worst case symbol rate of 5 Msps. This effect causes an expected E_b/N_0 degradation of no more than a few tenths of a dB.
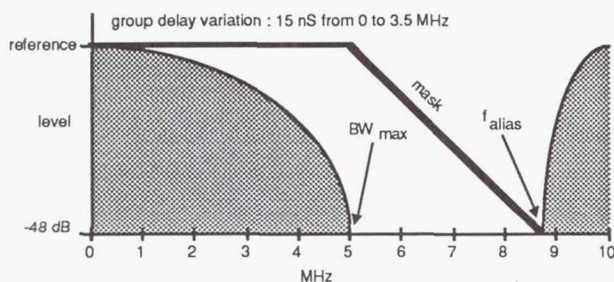

Figure 21 : Spectral Reconstruction Filter Mask

CONCLUSION

The architecture and hardware prototype of a flexible trellis-coded modem and codec transmitter has been presented. The FTMC transmitter uses digital modulation and coding, multi-symbol digital pulse shaping, and digital precompensation to tradeoff complexity for improved signal quality, bandwidth, and power efficiency. This type of architecture is envisioned to be of use in FDMA satellite uplinks and TDM or CDM satellite downlinks.

In any planned operational satellite network, the choice of modulation and coding techniques is dictated heavily by numerous other system concerns. However, offering future system designers workable flexible modems and codecs may relieve specifications in other subsystems and offer an enhanced satellite system that can offer more varied and higher quality services.

ACKNOWLEDGEMENT
The authors would like to thank L. Shimoda of Ohio University for generating the simulation data for Figures 8 through 11.

REFERENCES

[1] Viterbi, A. J., et al: A Pragmatic Approach to Trellis-Coded Modulation. IEEE Communications Magazine, July 1989, pp. 11-19.

[2] Harrold, J. L.; Budinger, J. M.; and Stevens,G. H.: On-Board Switching and Processing, Proc. IEEE, vol. 78, No. 7, July 1990, pp.1206-1212.

[3] Nuspl, P. P., et al: On-board Processing for Communications Satellites : Systems and Benefits, International Journal of Satellite Communications, Vol. 5, 1987, pp. 65-76.

[4] Campanella, S. J.: Communications Satellites: Orbiting into the 90's, IEEE Spectrum Magazine, August 1990, pp. 49-52.

[5] Hansen, J. W., et el: US TWTs from 1 to 100 GHz, Microwave Journal • 1989 State of the Art Reference, pp. 179- 193.

[6] Saha, D.; and Birdsall, T. G.: Quadrature-Quadrature Phase-Shift Keying, IEEE Transactions on Communications, Vol. 37, No. 5, May 1990, pp. 437- 448.

[7] Sklar, B.: Digital Communications Fundamentals and Applications, Prentice Hall, 1988.

[8] Lee, P. J.: Computation of the Bit Error Rate of Coherent M-ary PSK with Gray Code Bit Mapping, IEEE Transactions on Communication, Vol. 34, No. 5, May 1986, pp 488- 491.

[9] Poklemba, J. J: Programmable Digital Modem, NASA Space Communications Technology Conference, November 1991.

[10] Aghvami, A. H.; and Gemikonakli, O.: 16-ary QAM System for Low Data-Rate Satellite Services, Electron. Lett., Vol. 25, No. 16, August 3, 1989, pp. 1055- 1056.

[11] Allan, R. D.; Bramwell, J. R.; and Tomlinson, M.: A High Performance Satellite Data Modem Using Real-Time Digital Signal Processing. Olympus Utilisation Conference, Vienna, Italy, May 1989, pp. 51-55.

[12] Karam, G.; and Sari, H.: Analysis of Predistortion, Equalization, and ISI Cancellation Techniques in Digital Radio Systems with Nonlinear Transmit Amplifiers.

[13] Biglieri, E.; Barberis, S.; and Catena, M.: Analysis and Compensation of Nonlinearities in Digital Transmission Systems, IEEE Journal on Selected Areas in Communications, Vol. 6, No. 1, January 1988, pp. 42- 51.

[14] Cheung, S. W.; Aghvami, A.H.: Performance of a 16-ary DEQAM Modem Employing a baseband or RF Predistorter Over a Regenerative Satellite Link, IEE Proceedings, Vol. 135, No. 6, December 1988, pp. 547- 557.

[15] Karam, G.; and Sari, H.: A Data Predistortion Technique with Memory for QAM Radio Systems, IEEE Transactions on Communications, Vol. 39 , No. 2, February 1991, pp. 336- 344.

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1991	Conference Publication

4. TITLE AND SUBTITLE

Space Communications Technology Conference
Onboard Processing and Switching

6. AUTHOR(S)

5. FUNDING NUMBERS

WU-650-60-21

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135-3191

8. PERFORMING ORGANIZATION REPORT NUMBER

E-6548

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D.C. 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA CP-3132

11. SUPPLEMENTARY NOTES

Responsible person, James M. Budinger, (216) 433-3496.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Categories 17 and 32

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This proceedings of the NASA Space Communications Technology Conference held on November 12-14, 1991, in Cleveland, Ohio, presents onboard processing and switching technologies for application in future satellite communications systems. Onboard processing for B-ISDN (broadband integrated services digital network) services, packet-switched FDMA/TDM (frequency division multiple access/time division), a novel variation of CDMA (code division multiple access), and the merits of OBP (onboard processing) for low Earth orbiting mobile systems are addressed in the section, Satellite Network Architectures. Telecommunications services for the next generation of networks, a traffic analysis of B-ISDN, advanced lightwave networks, and survivable system processing are covered in Network Control and Protocols. Papers on bandwidth efficient coding, artificial intelligence for Earth terminals, multicarrier demodulation, neural-network-based video compression, an expert system for fault diagnosis, and laboratory measurements of INTELSAT OBP subsystems are presented in Concurrent Poster Presentations and Demonstrations. Two papers on expert systems for space hardware diagnostics and fault tolerance applied to fiber-optic networks and multichannel demultiplexers are covered in Fault Tolerance and Autonomy. Digital, optical, and acousto-optical approaches to multichannel demultiplexing and demodulation for onboard processed FDMA are presented in Multichannel Demultiplexing and Demodulation. In Information Switching and Routing, an onboard baseband switch, an OBP payload for asynchronous relay satellite networks, OBP for future telecommunications services, and congestion control for satellite packet switching are presented. In Modulation and Coding, the final section of this document, advanced modulation and coding technologies are applied to a B-ISDN modem-codec, a flexible, high-speed codec, and two variable-rate digital modems. Decoding multilevel demodulation and two approaches to digital pulse-shaped modulation for nonlinear satellite communications are also presented. Papers for the final two sessions of the conference were not available at the time of publication. These sessions, which addressed planned communications satellite systems under development at INTELSAT, NASA, Motorola, and Space Systems/Loral, and a panel session on the issues and challenges of onboard processing and switching technology insertion, are covered in a companion publication.

14. SUBJECT TERMS

Onboard processing; Modulation; Coding; Autonomy; Switching;
Routing; Network control

15. NUMBER OF PAGES

312

16. PRICE CODE

A14

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	